NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

A METHODOLOGY FOR EVALUATING
THE RELATIONSHIP BETWEEN MEASURES OF
EVALUATION (MOEVs): THE STF APPROACH

by

Stephen Charles Kessner

June, 1991

Thesis Advisor: Donald A. Lacer

Approved for public release; distribution is unlimited

T258466

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b. OFFICE SYMBOL (*If applicable*) 39	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c. ADDRESS (*City, State, and ZIP Code*) Monterey, CA 93943-5000	7b. ADDRESS (*City, State, and ZIP Code*) Monterey, CA 93943-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (*If applicable*)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (*City, State, and ZIP Code*)	10. SOURCE OF FUNDING NUMBERS Program Element No / Project No / Task No / Work Unit Accession Number

11. TITLE (*Include Security Classification*)
A METHODOLOGY FOR EVALUATING THE RELATIONSHIP BETWEEN MEASURES OF EVALUATION (MOEVs): THE STF APPROACH

12. PERSONAL AUTHOR(S) Kessner, Stephen C.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (*year, month, day*)	15. PAGE COUNT
Master's Thesis	From To	June 1991	147

16. SUPPLEMENTARY NOTATION
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

17. COSATI CODES			18. SUBJECT TERMS (*continue on reverse if necessary and identify by block number*)
FIELD	GROUP	SUBGROUP	Command and Control (C^2), Measure of Effectiveness (MOE), Measure of Evaluation (MOEV), Measure of Force Effectiveness (MOFE), Measure of Merit (MOM), Measure of Performance (MOP), Measure of Platform Effectiveness (MOPE), Measure of Success (MOS)

19. ABSTRACT (*continue on reverse if necessary and identify by block number*)

This thesis shows how the subjective transfer function (STF) approach can be used to evaluate the relationship between measures of evaluation (MOEVs). This methodology, developed by The Rand Corporation, uses algebraic modeling and expert opinion to measure the effectiveness of complex systems.

Measures of evaluation (MOEVs) are defined and a relationship between the MOEVs is developed. This hierarchy of MOEVs is derived from a defense planning process and a corresponding military system structure. The STF approach is described in detail. An example is provided showing the relationship between the measure of effectiveness (MOE) of tactical air command and control and the measure of force effectiveness (MOFE) of tactical air forces. Finally, the STF approach is evaluated from two perspectives. First, it is evaluated as an evaluation methodology, using Hollnagel's six criteria. Second, Giordano and Weir's criteria for evaluating a model's effectiveness and construction are discussed.

Recommendations are made in the following areas: 1) the defense planning structure, 2) the military system hierarchy, 3) the definintions of MOEVs, 4) the hierarchical relationship among MOEVs, and 5) the assessment of evaluation methodologies.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS REPORT ☐ DTIC USERS	UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Donald A. Lacer	22b. TELEPHONE (*Include Area code*) (408) 646-3446	22c. OFFICE SYMBOL CC/LA

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

Approved for public release; distribution is unlimited.

A Methodology for Evaluating
The Relationship Between Measures of Evaluation (MOEVs):
The STF Approach
by

Stephen Charles Kessner
Captain, United States Air Force
B.S., Norwich University, 1983

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN SYSTEMS TECHNOLOGY
(Command, Control, and Communications)

from the

ABSTRACT

This thesis shows how the subjective transfer function (STF) approach can be used to evaluate the relationship between measures of evaluation (MOEVs). This methodology, developed by The Rand Corporation, uses algebraic modeling and expert opinion to measure the effectiveness of complex systems.

Measures of evaluation (MOEVs) are defined and a relationship between the MOEVs is developed. This hierarchy of MOEVs is derived from a defense planning process and a corresponding military system structure. The STF approach is described in detail. An example is provided showing the relationship between the measure of effectiveness (MOE) of tactical air command and control and the measure of force effectiveness (MOFE) of tactical air forces. Finally, the STF approach is evaluated from two perspectives. First, it is evaluated as an evaluation methodology, using Hollnagel's six criteria. Second, Giordano and Weir's criteria for evaluating a model's effectiveness and construction are discussed.

Key issues and recommendations are made in the following areas: 1) the defense planning structure, 2) the military system hierarchy, 3) the definitions of MOEVs, 4) the hierarchical relationship among MOEVs, and 5) the assessment of system evaluation methodologies.

TK5101
K34477
c.1

TABLE OF CONTENTS

I. INTRODUCTION 1

A. BACKGROUND. 1

B. SCOPE. 2

II. MEASURES OF EVALUATION (MOEVs). 4

A. OVERVIEW 4

B. MEASURES OF EVALUATION (MOEVs) 4

C. HIERARCHY OF MOEVs 7

1. OVERVIEW 7

2. DEFENSE PLANNING FRAMEWORK 7

3. RELATIONSHIP OF MOEVs TO THE DEFENSE PLANNING FRAMEWORK 10

III. TACTICAL AIR FORCES AND TACTICAL AIR C^2 14

A. OVERVIEW 14

B. THE AIR/LAND BATTLE AND THE TACTICAL AIR FORCE 14

C. AIR/LAND C^2 SYSTEMS AND TACTICAL AIR C^2 SYSTEMS 16

D. ELEMENTS OF THE TACTICAL AIR FORCE 18

E. ELEMENTS OF TACTICAL AIR C^2 23

F. MOFEs FOR THE TACTICAL AIR FORCE 29

G. MOEs FOR TACTICAL AIR C^2 30

IV. A METHODOLOGY FOR RELATING MOEs TO MOFEs 33

A. OVERVIEW 33

B. THE SUBJECTIVE TRANSFER FUNCTION (STF) APPROACH . . . 33

1. Defining a Complex System Representation 36

a. Structural Hypotheses 37

(1) Identifying System Outcomes and Factors. 37

(2) Identifying Alternative Structures. 42

b. STF Hypotheses 45

2. Conducting Judgement Experiments 50

a. Designing Experiments to Test Hypotheses 51

(1) Factorial Designs: Tests Between Interactive and Noninteractive Functions. 51

(2) Factorial Designs: Extensions Which Permit Additional Tests of Hypotheses. 56

b. Selecting Experimental Designs 68

(1) Pilot Study Phase. 68

(2) The STF Testing Phase. 70

c. Collecting the Judgements Data 72

3. Determine the STFs and the Final Structure 73

a. Data Analysis 73

(1) Determine Individual Differences. 73

(2) Testing Between the STFs. 74

b. The Final System Structure: An Example 79

4. Evaluate Systems 81

V. APPLICATION OF THE STF APPROACH TO THE EVALUATION OF TACTICAL AIR C^2 90

A. OVERVIEW 90

B. DEFINING A COMPLEX SYSTEM REPRESENTATION FOR TACTICAL AIR C^2 AND TACTICAL AIR FORCES 91

1. Structural Hypotheses 91

a. System Outcomes and System factors 92

b. Alternative Structures 93

2. STF Hypotheses 101

C. CONDUCTING THE JUDGEMENT EXPERIMENTS 101

1. Designing the Experiments 101

2. Constructing Definitions of System Outcomes and System Factors. 105

D. EVALUATING TACTICAL AIR C2 AND TACTICAL AIR FORCES 107

VI. EVALUATION CRITERIA FOR ASSESSING THE STF APPROACH . . . 112

A. OVERVIEW 112

B. EVALUATING THE APPLICATION OF THE STF APPROACH METHODOLOGY 112

1. Evaluation of the STF Approach Methodology 113

2. Evaluation of Algebraic Modeling of Expert Responses 120

a. Evaluation of the Effectiveness of a Model 121

b. Evaluation of the Construction of a Model 122

VII. SUMMARY AND RECOMMENDATIONS 125

A. OVERVIEW 125

B. SUMMARY 125

C. KEY ISSUES AND RECOMMENDATIONS 128

1. Issue No. 1: The Defense Planning Framework 128

2. Issue No. 2: The Military System Hierarchy 128

3. Recommendation No. 1: Standardize Definitions of MOEVs 129

4. Recommendation No. 2: Standardize MOEV Hierarchy 130

5. Recommendation No. 3: Develop A Standard for Assessing Evaluation Methodologies 130

LIST OF REFERENCES 131

INITIAL DISTRIBUTION LIST 133

LIST OF TABLES

TABLE I. DEFINITIONS OF FACTORS AND OUTCOMES FOR IMMEDIATE TARGETING STRUCTURE SHOWN IN FIGURES 12 AND 13. . . 40

TABLE II. POSSIBLE STFs. 49

TABLE III. PRELIMINARY HYPOTHESES ASSOCIATED WITH COMPONENTS SHOWN IN FIGURES 27 AND 29. 95

TABLE IV. OUTLINE OF A POSSIBLE JUDGEMENT EXPERIMENT FOR UNIT 6 AT THE FUNCTION LEVEL. 103

TABLE V. OUTLINE OF POSSIBLE JUDGEMENT EXPERIMENT AT THE TAO TIER. 104

TABLE VI. SUMMARY OF ASSESSMENT FACTORS. 117

TABLE VII. SUMMARY OF OVERALL ASSESSMENT OF EVALUATION CRITERIA. 119

LIST OF FIGURES

Figure 1A. Force planning structure. 9

Figure 1B. Military System Hierarchy. 9

Figure 2A. Relationship of the Military System to the Defense Planning Process. 11

Figure 2B. Hierarchy of MOEVs. 13

Figure 3. Air/Land Battle System. 15

Figure 4. The Mix of Functional Systems. 17

Figure 5. The Mix of Functional Systems and C^2. 17

Figure 6. A Tactical Air Force Structure. 19

Figure 7. Tactical Air Resource Employment. 20

Figure 8. A Tactical Air Force Structure: Tactical Air Resources. 21

Figure 9. A Tactical Air Force Structure: Tactical Air Operations. 24

Figure 10. A Tactical Air Force Structure: Assuming an Invariant Decision Making Process.[Ref. 8:p.4,6] 25

Figure 11. Tactical Air Command and Control.[Ref. 6:p.4] 27

Figure 12. Hypothesized Immediate Targeting Structure. 38

Figure 13. Alternative Structure to Figure 12. 44

Figure 14. Final Structure of a C^2 System. 46

Figure 15. Outline of Subjective Measurement. 47

Figure 16. Possible Hypothetical Data Obtained From Two-way Factorial Designs. 52

Figure 17. Possible Hypothetical Data Obtained From Two-way Factorial Designs. 54

Figure 18. Experimental Design (for three factors) Which Varies the Amount of Information. 57

Figure 19. Predictions of Four Different Functions When Designs Vary the Amount of Information. 60

Figure 20. Graphic Predictions of Additive and Averaging Functions. 64

Figure 21. Hypothetical Data Predicted by Different Algebraic Functions. 76

Figure 22. Alternative Structure/Function Combination. 77

Figure 23. The Final Structure for Immediate Targeting With Corresponding STFs. 80

Figure 24A. System Comparisons. 83

Figure 24B. System Comparisons. 84

Figure 24C. System Comparisons. 85

Figure 25. Subjective Scale Value for the Location/Classification, Coverage, and Currency Factors Shown in Figure 12. 86

Figure 26. Theoretical Predictions. 88

Figure 27. Tactical Air C2 and Tactical Air Force Representation. 94

Figure 28. Tactical Air C^2 and Tactical Air Force Representation: Experimental Units. 96

Figure 29A. Alternative Tactical Air C^2 and Tactical Air Force Representation: Fixed Targets. 98

Figure 29B. Alternative Tactical Air C^2 and Tactical Air Force Representation: Stationary Force Elements. 99

Figure 29C. Alternative Tactical Air C^2 and Tactical Air Force Representation: Moving Force Elements. 100

Figure 30. Tactical Air C^2 and Tactical Air Force Representation: Transfer Functions. 108

Figure 31. Evaluation of Tactical Air C^2 and Tactical Air Force Representation. 110

I. INTRODUCTION

A. BACKGROUND.

As man continues to increase his control of the environment surrounding him, it requires an in depth knowledge of the elements of which it is composed, and an evaluation of the effect each one has on the whole entity - the environment. Similarly, it is no different in controlling combat effectiveness. Man strives to strengthen his advantage over an enemy or threat by making improvements to the elements he has control over: tactics, technology, and training. By evaluating his effectiveness in each of these areas, he can determine which are deficient and require improvement. Manpower and money can then be wisely invested in these areas, resulting in an improvement of man's overall war fighting capability.

In evaluating each element, measures of evaluations (MOEVs) or measures of assessment have been developed. The following measures are defined in Chapter II: measures of merit (MOM), measures of performance (MOP), measures of effectiveness (MOE), measure of platform effectiveness (MOPE), measures of force effectiveness (MOFE), and measures of success (MOS) [Ref. 1:p.1_10]. These measures are not all inclusive or standardized. However, in the opinion of the author, they are the MOEVs commonly used when analyzing system effectiveness. [Ref. 2:p.299].

A problem arises in determining how to evaluate the impact which the effectiveness of each element contributes to the effectiveness of the whole system. A number of approaches exist, yet it continues to be an area which challenges experts in determining how to quantitatively assess the contribution which a system makes to the overall war fighting capability. C^2 has been called a "force multiplier"; however, a more appropriate nickname might be "force mystifier". For the present, it is the author's opinion, that no one knows how to determine the multiplier in exacting detail. However, several models or approaches exist which allow us to analyze, in general detail, the effect of C^2.

B. SCOPE.

This thesis will look at one methodology for evaluating the relationship between MOEVs: the subjective transfer function (STF) approach. It will be used to show how the effectiveness of tactical air C^2 and the impact it has on the force effectiveness of tactical air forces can be evaluated. Chapter II will define the MOEVs and describe the hierarchy of MOEVs used in this thesis. Chapter III will provide background information on the structure of tactical air forces and tactical air C^2, the major elements of each, and their relationships. The MOFE and MOE which will be used in evaluating tactical air forces and tactical air C^2, respectively, will be defined. Chapter IV will describe the STF approach methodology, developed by Monti Callero and Clairice T. Veit, both of The Rand Corporation. Chapter V will show how the STF approach could be applied to evaluate the relationship between the MOE of tactical air C^2 and the MOFE of tactical air forces. Chapter VI will look at criteria for assessing the STF approach. Erik

Hollnagel's six criteria for evaluating expert systems will be shown how it can be used to evaluate the STF approach methodology. Frank R. Giordano and Maurice D. Weir's criteria for evaluating models will be discussed for evaluating the algebraic modeling aspect of the STF approach. Finally, in Chapter VII, the main points of this thesis will be summarized. Recommendations will be made on what needs to be done, before the credibility of a methodology for relating MOEVs, such as the STF approach can be verified.

II. MEASURES OF EVALUATION (MOEVs).

A. OVERVIEW

According to the Merriam-Webster Dictionary, "evaluate" is synonymous with "appraise" and "value". A MOEV as used in this thesis is a unit of measurement which represents the value of a system or the elements which comprise the system. Since a system can be decomposed into different levels, a series of MOEVs have been developed which correspond to the different levels of a system.

This thesis considers the United States Air Force (USAF) tactical air forces as the system being evaluated. It will look at one of the elements of tactical air forces, namely, tactical air C^2. Since the objective of this thesis is to look at one methodology for evaluating the relationship between MOEVs, it is important to define the MOEVs being used and the hierarchy defining the relationship between them.

This chapter will define the MOEVs used in this thesis and describe the hierarchy of the MOEVs.

B. MEASURES OF EVALUATION (MOEVs)

A measure is a variable used in the appraisal of a process [Ref. 3:p.16]. It is important that a set of comprehensive measures are defined and used, which capture all of the system elements that affect the overall system performance. A well defined measure can either be qualitative or quantitative in nature.

The focus of this thesis is not on defining new measures, but on the interaction between the measures of a part of the system, or system element on the whole system. However, it is important to ensure all relevant aspects of a system are captured by the MOEVs.

The following is a glossary of the relevant MOEVs, which are vital to capturing the key variables in a system. They are adapted from an informal document produced by SPAWAR 317, entitled, *Suggested Assessment Measure Definitions/Structure* [Ref. 1:p.1_10].

1. **MEASURE OF MERIT (MOM).** An assessment measure which addresses the non-operational attributes, which are relevant to decision factors such as cost, risk, and availability.

2. **MEASURE OF PERFORMANCE (MOP).** An assessment measure of the degree to which hardware and software in the system perform with respect to established system specifications, independent of operator actions [Ref. 1:p.3_4].

3. **MEASURE OF EFFECTIVENESS (MOE).** An assessment of how well the system makes or assists the commander in making decisions for any and all missions [Ref. 1:p. 3_3].

4. **MEASURE OF PLATFORM EFFECTIVENESS (MOPE).** An assessment measure of the all the systems of a platform or a unit.

5. **MEASURE OF FORCE EFFECTIVENESS (MOFE).** An assessment measure of the relative value of different force compositions for accomplishing multiple missions.

6. **MEASURE OF SUCCESS (MOS).** An assessment measure of the relative value of the overall system. It is used to state requirements or to assess and compare architectural options. A wide variety of approaches have been employed for assessing the attributes of a system and its elements. A report done for the Navy's Space and Naval Warfare System Command [Ref. 1] came up with the following conclusion on force and unit measures.

> No methodology has been devised to translate TLWR [top level warfare requirements] statements into force and unit level performance criteria. As a result, the performance criteria for the elements being produced by the Systems Commands are not founded upon a common requirements base. When developed and integrated there is no assurance that the elements will collectively meet unit and force level performance criteria. [Ref. 1:p. 7-2]

Even a standardized set of terminology has not been realized for MOEVs like MOP and MOE.

> The terminology used in MOPs and MOEs varies significantly from user to user. Document searches became meaningless when the terms were defined so different as to be complete opposites. Many of the documents defined MOEs in terms of this study ... are described as MOPs. [Ref 1: p. 7_1]

The above definitions of MOEVs were chosen because they appeared to be the most comprehensive and straight forward in evaluating all the key components of a military system. The next section will show how relating the MOEVs to a military system is important in defining a hierarchical relationship between MOEVs.

C. HIERARCHY OF MOEVs

1. OVERVIEW

Before defining the hierarchy of MOEVs to be used it is important to look at the relationship of MOEVs to the system they are evaluating. MOEVs assist in verifying that the system design meets stated requirements. It is necessary to understand the goals and objectives of a system, if we are to choose appropriate and useful MOEVs to evaluate it.

2. DEFENSE PLANNING FRAMEWORK

A Rand report prepared for the United States Air Force (USAF) and the Office of the Secretary of Defense, in part by the National Defense Research Institute, presents an approach to strengthening the U.S. defense planning process [Ref. 4:p.iii]. The report, *A Framework for Defense Planning*, relates official statements of national security, national military strategy, and the operational capabilities of force elements, to programs for developing and procuring military systems and services [Ref. 4:p. iii]. This framework will be used to provide a hierarchy for a systematic and operationally oriented process, which determines whether or not a need exists for a particular system. This approach was chosen because it is operationally oriented and can be applied to all branches of the armed forces, not only to the USAF, encouraging "jointness" among the services. Within the defense planning framework, we are going to be interested in force planning, and relating the MOEVs to the various levels of force

planning. The following quote describes the key elements of force planning (note italics were added).

> Force planning (or development), an aspect of defense planning, includes organizing, equipping, training, upgrading, maintaining, and supporting various *force elements* to provide specific *operational capabilities*. The force element, an organizational unit, consists of the personnel and major items of equipment-tanks, ships, aircraft-together with all the supporting resources needed to provide it with a stated operational capability. The specific operational capabilities provided by a specific force element is the sum of the *operational tasks*, or individual military operations, which that force element is capable of performing. The capability is specific with respect to the type of tasks performed, as well as to the magnitude of the effort over time. Operational tasks are the building blocks of operational capability. [Ref. 4:p. 1]

A top-down look at this force planning structure (Figure 1A) starts with the national security strategy, which is made up of the nation's economic, political, and intellectual power in conjunction with the national military strategy. The national military strategy consists of regional strategies, which are underwritten by a "cluster of *operational objectives*". Finally, the means of achieving the stated operational objectives is determined by our capability to accomplish "a cluster of operational tasks". [Ref. 4:p. v]

A top-down look at the military system (Figure 1B) can be thought of as a hierarchy composed of the following levels of elements: national military strategy, force, platform/unit, system, and equipment. If we want to evaluate the effectiveness/ performance of a tactical air C^2 system with respect to the military system, we must go one level higher and include national military strategy as the top level in the military system. For it is this strategy which drives the requirements/needs for the rest of the military system. The key to the force planning approach of relating strategies to tasks is

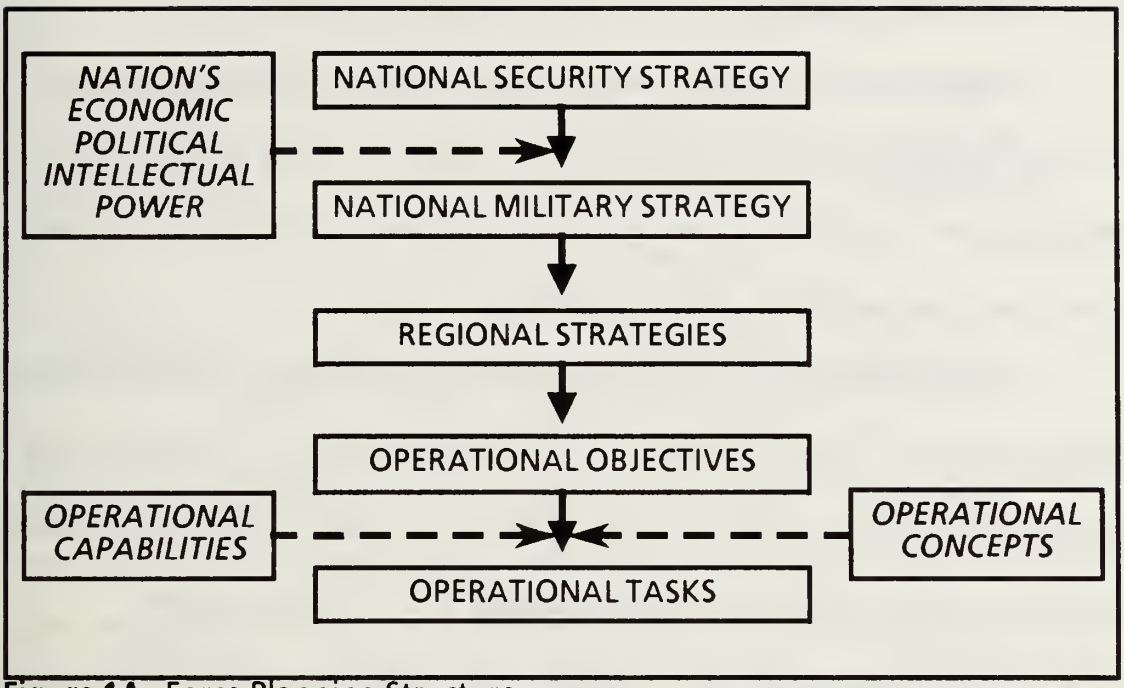

Figure 1A. Force Planning Structure.

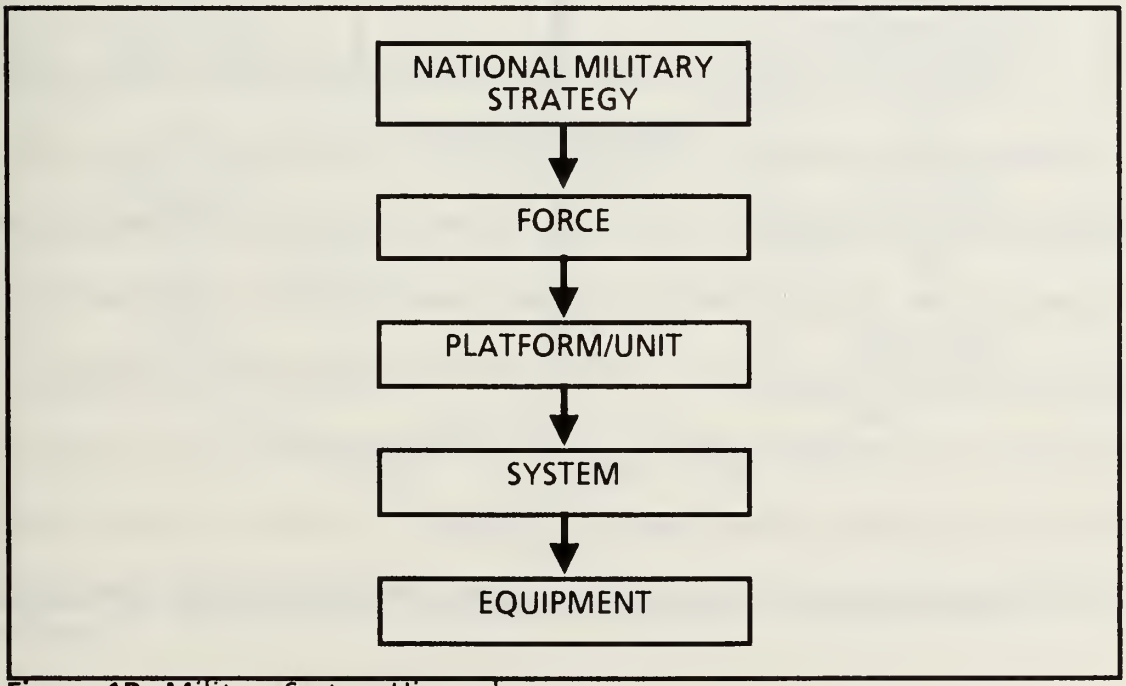

Figure 1B. Military System Hierarchy.

the use of operational objectives as a means of defining operational capabilities that a Service is expected to provide.

Operational capabilities describe what operational tasks relating to the operational objective can be accomplished at what rate during what period. By identifying these capabilities we can assess the current and future ability of our force to perform operational objectives. [Ref. 4:p. 16]

Operational concepts define the means of accomplishing operational tasks and achieving operational objectives, as well as defining the programs to be implemented to provide the equipment to accomplish the operational concept. In general, operational concepts show how particular systems are going to be acquired and allocated.

3. RELATIONSHIP OF MOEVs TO THE DEFENSE PLANNING FRAMEWORK

Now that the defense planning concept has been described, we can establish a relationship which could exist between the hierarchy for defense planning and a hierarchy for MOEVs. Figure 2A provides a suggested methodology for looking at the relationship between Rand's defense planning structure and MOEVs, by adapting a similar approach used for evaluating the Tactical Command System (TCS) measures of effectiveness and performance. [Ref. 1:p. 1_9] The left column of Figure 2A shows the hierarchy of the defense planning structure (Figure 1A). It shows the key levels within the process, which determine what systems (i.e. tactical C^2) are required and appropriated

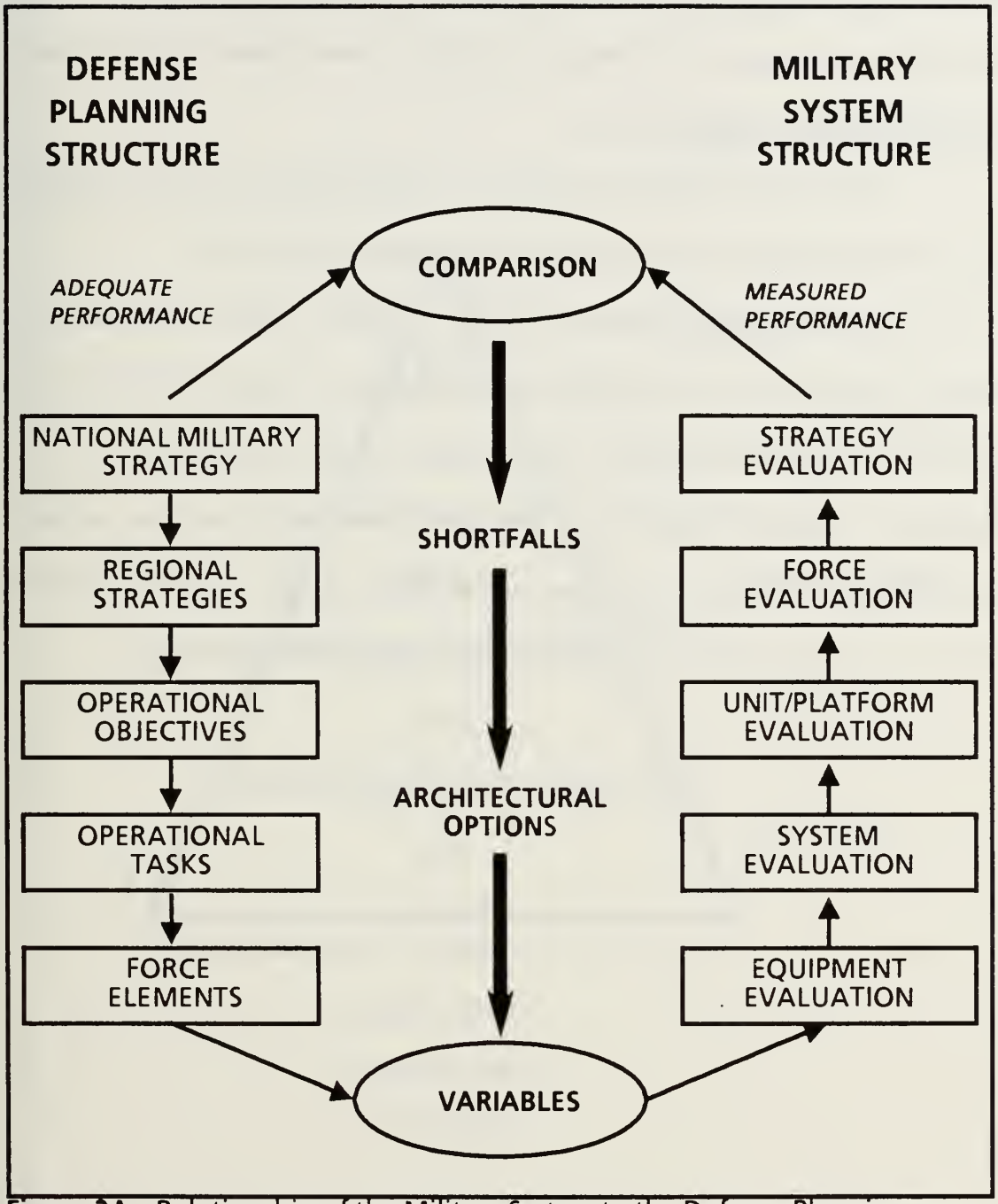

Figure 2A. Relationship of the Military System to the Defense Planning Process.

by the Department of Defense. Each level determines both adequate or required performance characteristics of the system. The second column from the right of Figure 2A, shows the relationship of the levels of a military system (Figure 1B), which can be related to the defense planning process.

A hierarchy of MOEVs is depicted in Figure 2B. [Ref. 1:p. 1_12] Again the relationship between the MOEVs and the military system can be seen.

The relationship between each level of MOEVs will be examined in the following chapters. Without an established methodology for relating MOE of a system to the MOFE of a force we cannot determine the impact that system has on force effectiveness. Chapter IV looks at one methodology for relating the levels within the MOEV hierarchy. First, two components of the military system, tactical air forces and tactical air C2 and their respective MOEVs will be described in Chapter III.

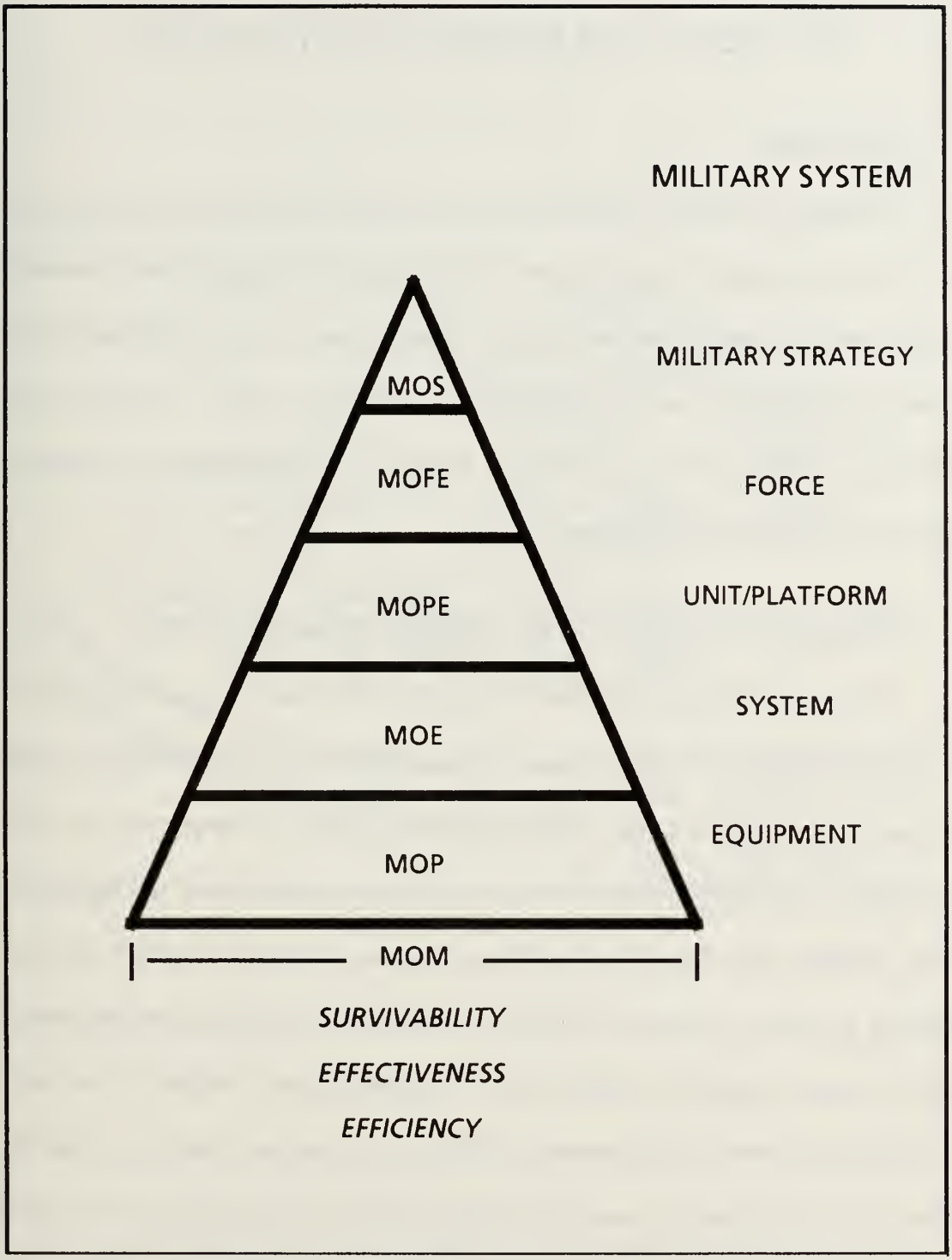

Figure 2B. Hierarchy of MOEVs.

III. TACTICAL AIR FORCES AND TACTICAL AIR C^2

A. OVERVIEW

This chapter provides a general description of the tactical air force and the tactical air C^2. The relationship between the two will be defined by looking at their relationships to the air/land battle and the air/land C^2 systems respectively. Next, the elements of both the tactical air force and tactical air C^2 will be described. Finally, a measure of force effectiveness for the tactical air force and a measure of effectiveness for tactical air command and control will be defined.

B. THE AIR/LAND BATTLE AND THE TACTICAL AIR FORCE

The air/land battle system can be thought of, in basic terms as an army corp consisting of divisions, brigades, battalions, etc. arrayed on the battlefield along with its accompanying tactical air force. Figure 3 provides a look at the structure of the air/land battle system. This portrayal shows only U.S. forces, excluding Navy and Marine Corp forces. However, the air/land battle system is far more complex than what has just been described and what is depicted in Figure 3. It includes multiservice and multinational commands and is fought by "theater forces". Theater forces are defined to be the land and tactical air forces, which together would fight the air/land battle in a land mass theater (i.e. North Atlantic Treaty Organization (NATO) Center, Korea, or the Middle East and its contained and adjacent waters). Note this differs from the definition

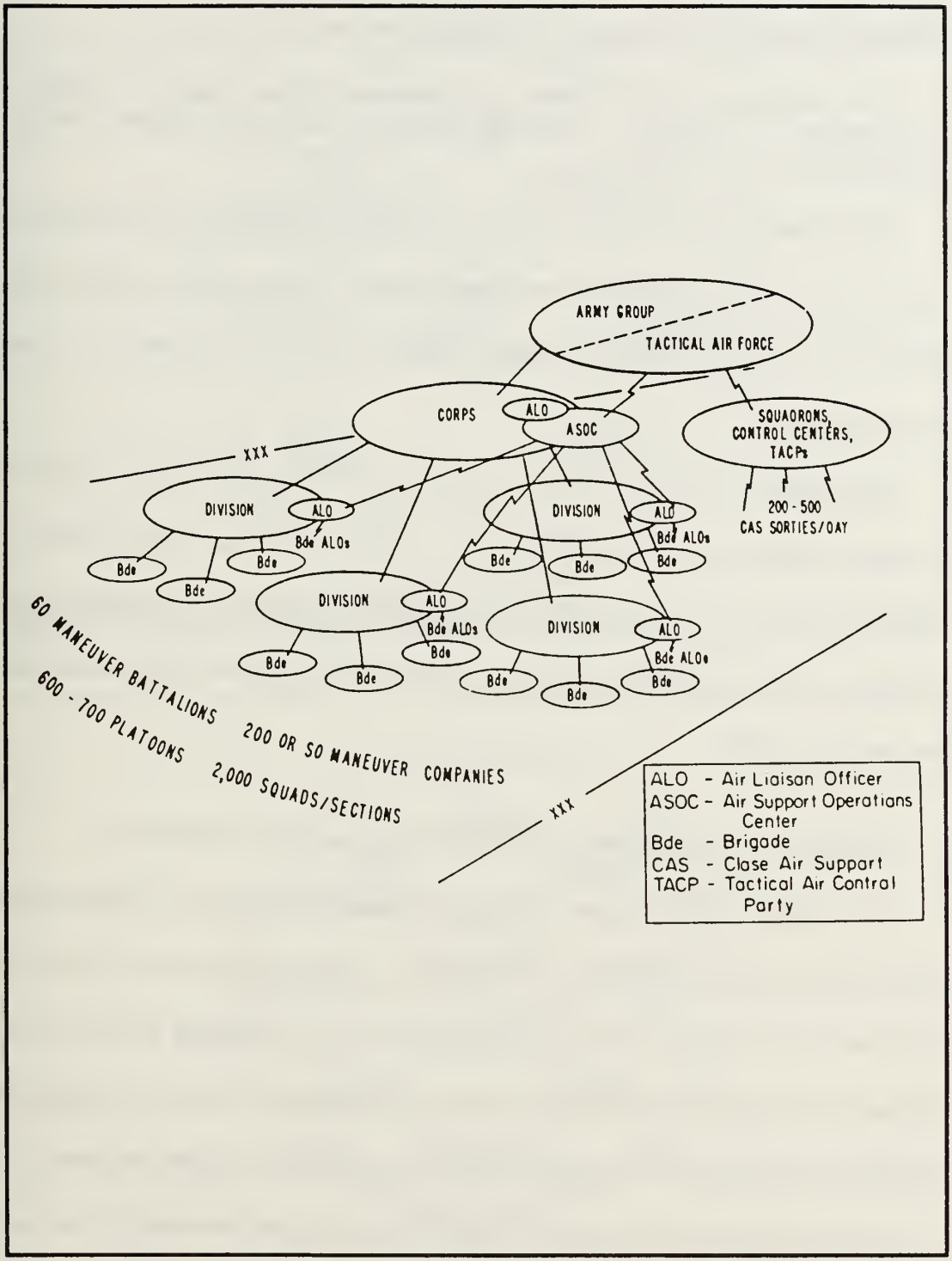

Figure 3. Air/Land Battle System.[Ref. 5:p. 6]

of "theater" according to the Department of Defense Dictionary of Military and Associated Terms, JCS Publication 1. It defines "theater" as

> 'The geographical area outside the continental United States for which a commander of a unified or specified command has been assigned military responsibility.' [Ref. 5:p.5]

The air/land battle system is a mixture of functional systems or subsystems, which occur at each level-corps, division, and brigade down to battalion and company. Figure 4 shows the mix of functional systems, which comprise each level to a certain degree.[Ref. 5:p.6]

The tactical air systems sector of Figure 4, which at the theater level is the tactical air forces, expands into all the other sectors. It is this part of the air/land battle system which this thesis looks at, in particular its tactical air C^2 element. As illustrated in Figure 5, the lines of communications cross over all the sectors and meet in the center forming command and control.

C. AIR/LAND C^2 SYSTEMS AND TACTICAL AIR C^2 SYSTEMS

The airland battle C^2 systems support an intricate combination of requirements for a land battle and a tactical air battle. Under the overall air/land commands, the direction and control of the tactical air and land battles are united through air liaison officers (ALOs) at the brigade, division, and corps levels; through the tactical air control parties (TACPs) at the battalion and brigade levels and the air support operations center (ASOC) at corps level; and through the close teamwork and mutual understanding of air and land commanders at each echelon.[Ref. 5:p.89]

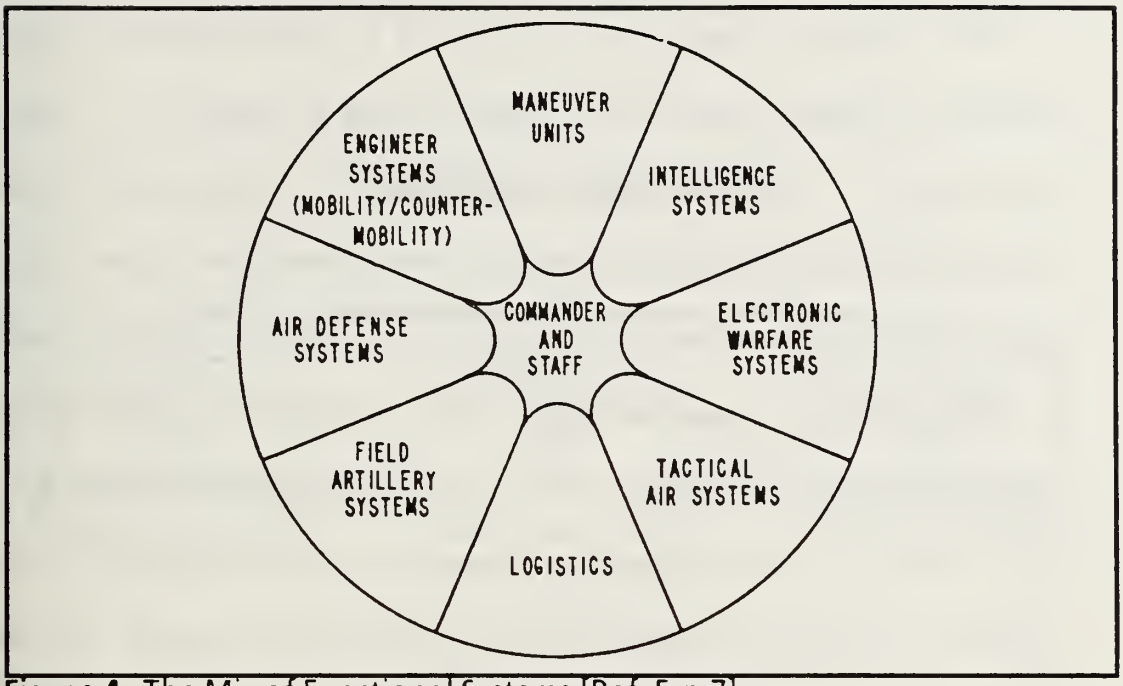

Figure 4. The Mix of Functional Systems.[Ref. 5:p.7]

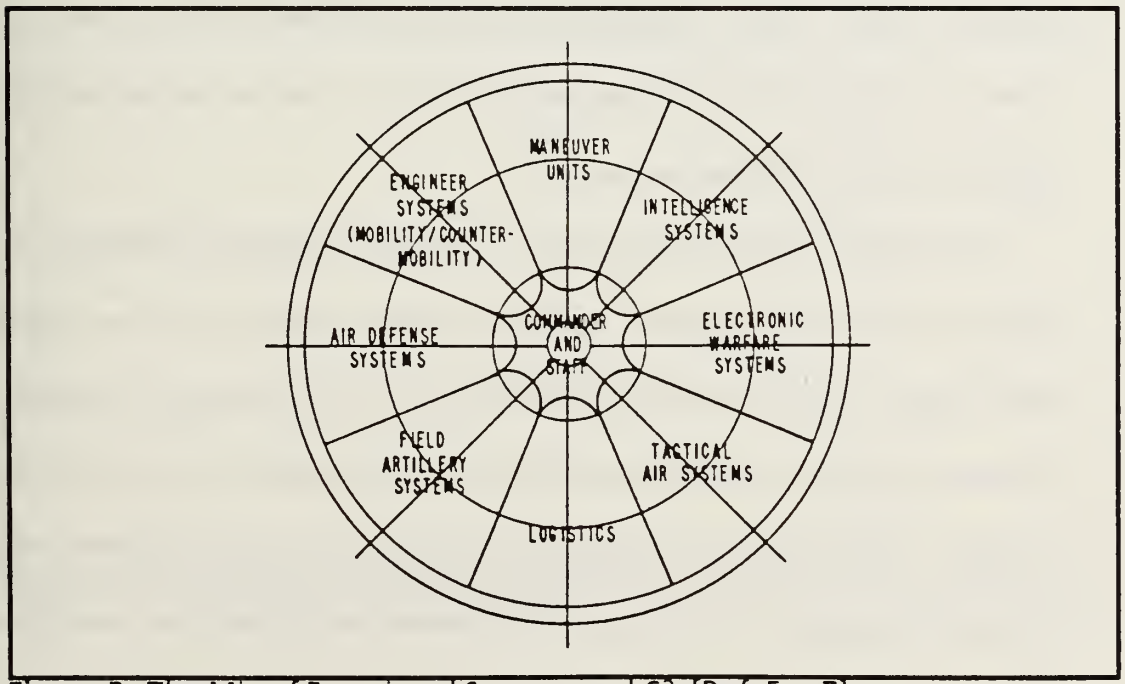

Figure 5. The Mix of Functional Systems and C^2. [Ref. 5:p.7]

D. ELEMENTS OF THE TACTICAL AIR FORCE

One definition of tactical air C^2 is that it is "...the means by which an air commander brings tactical air forces to bear against an enemy in war"[Ref. 6:p.1]. Before looking at tactical air C^2, it is important to state the other elements besides tactical air C^2, which make up tactical air forces. Tactical air forces can be divided into three areas (Figure 6): 1) tactical air resources, 2) tactical air C^2, and 3) tactical air operations (TAOs). The employment of tactical air forces is actually the employment of tactical air resources. At the higher levels within the military system hierarchy, TAO requirements are determined. It is the tactical air C^2 component of tactical air forces which spans both the TAO and tactical resources. It bridges and balances the two, by managing the resources and applying the tactical air force to combat needs.[Ref. 6:p.2] Figure 7 illustrates this relationship between the two areas of tactical air forces in the employment of tactical air resources. The next section will discuss the components which comprise tactical air C^2.

The tactical air force will be based on a typical USAF tactical air force. It consists of fighter aircraft, reconnaissance aircraft and transport aircraft, which are organized by tactical wings. Each tactical wing consists of 36 to 72 of one type of aircraft and the personnel, equipment, supplies, and facilities needed to maintain and operate those aircraft in combat. These items can be thought of as tactical air resources as shown in Figure 8. In addition, the tactical air force contains a C^2 system, which from a top-down perspective

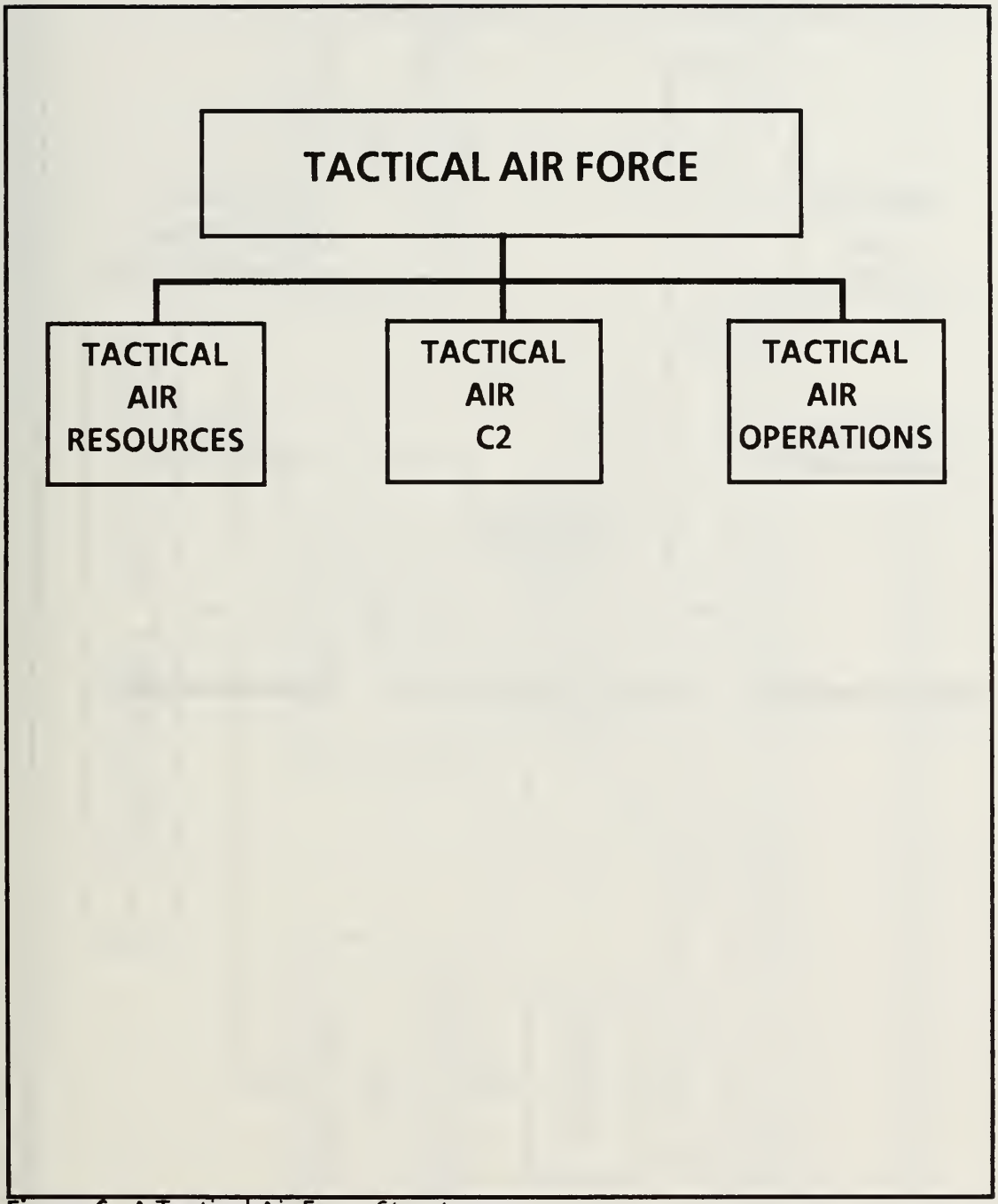

Figure 6. A Tactical Air Force Structure.

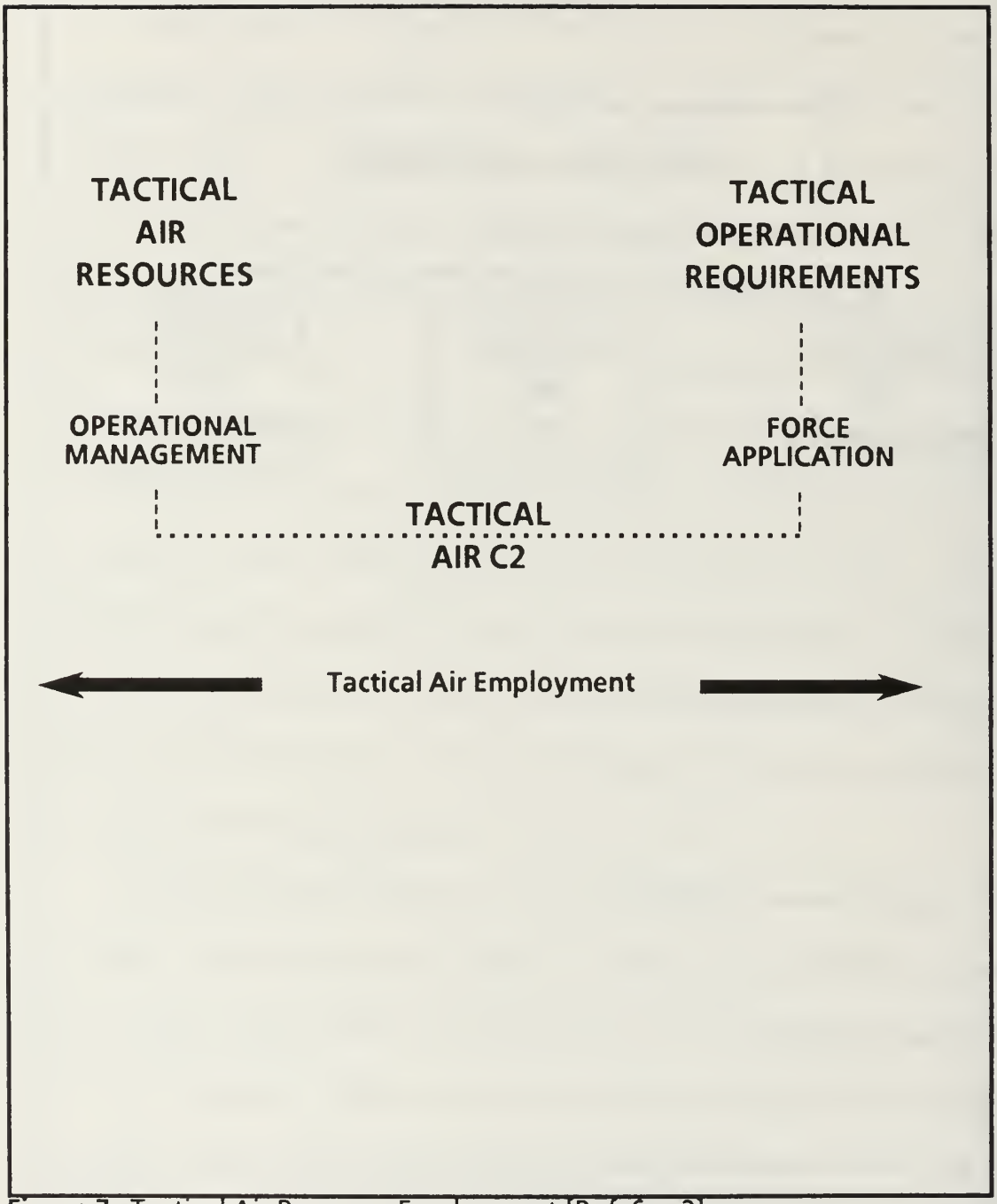

Figure 7. Tactical Air Resource Employment.[Ref. 6:p.2]

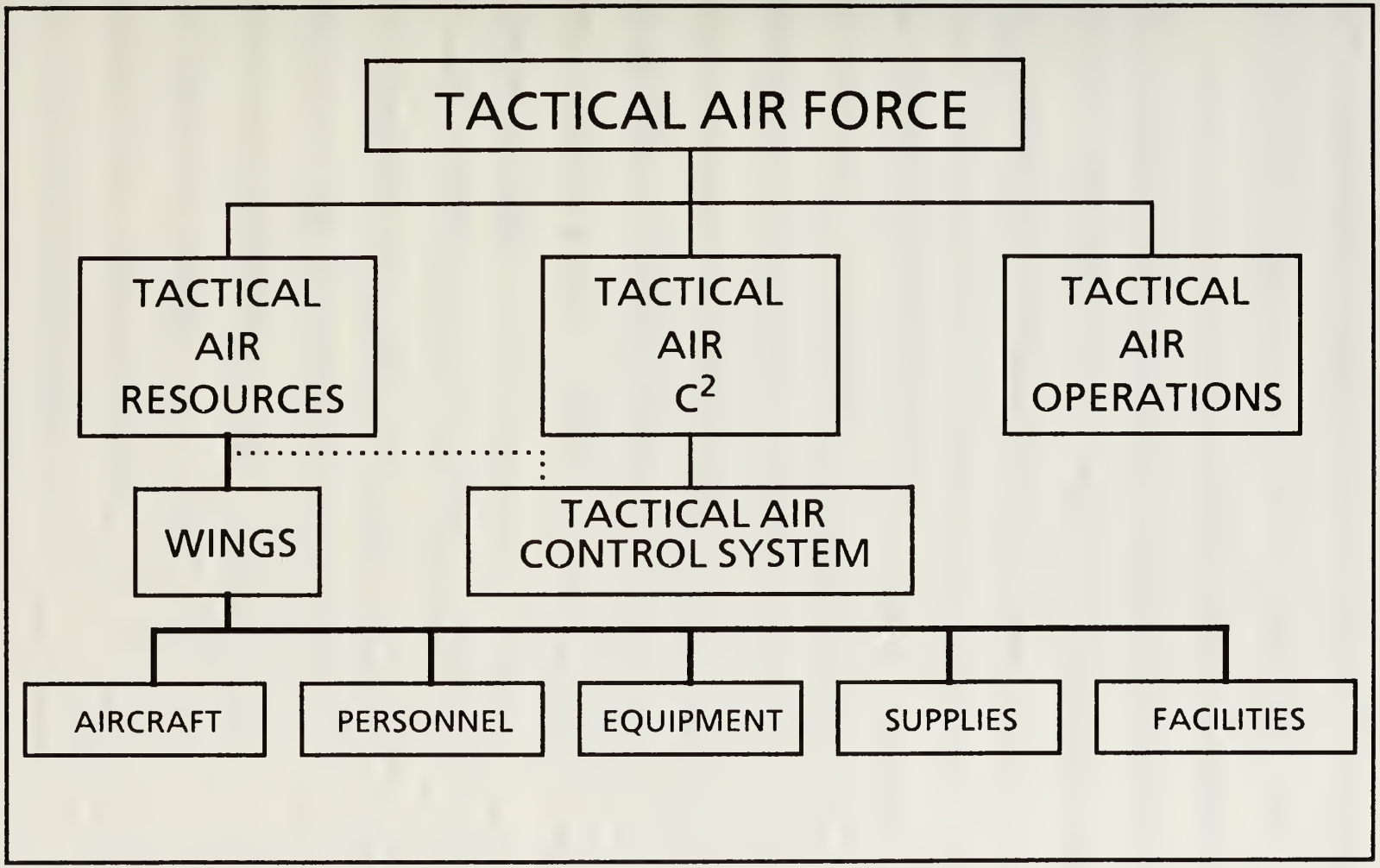

Figure 8. A Tactical Air Force Structure: Tactical Air Resources.

goes from the overall commander of the tactical air forces to the commanders of each of the wings. This system is called the Tactical Air Control System (TACS) and will be discussed in the next section.[Ref. 7:p.27]

The course of military events is affected by tactical air forces by flying or having the potential to fly combat missions. These missions are categorized in Tactical Air Operations (TAOs) and represent primary mission objectives. The following are TAOs: Air Defense, Reconnaissance, Search and Rescue, Offensive Counter Air, Close Air Support, Interdiction, and Airlift.[Ref. 7:p.27]A land battle can be defined as a single military event in which tactical air forces play an important role. An example of this is large-scale conventional warfare. Here opposing forces engage in a lot of ground battles in order to achieve military objectives (such as the occupation of territory or destruction of the opponent's forces), which are expected to contribute to the overall attainment of national goals. During these battles, while the army is engaging the enemy army forces on the ground, the tactical air forces conduct tactical air operations to influence the outcome of the battle. They perform *close air support* operations, by attacking and destroying enemy army forces, which are in direct contact with our army forces. They perform *airlift* operations, by flying in reinforcements and resupplies to our army forces. They perform *interdiction* operations, by attacking and destroying enemy forces and equipment and by closing roads and other lines of communication in the enemy's rear to keep new enemy forces from joining the battle. They perform *air defense* operations, by attacking and destroying enemy aircraft attempting to attack our army forces. Finally,

they perform the other TAOs, which have less of a direct influence on the course of the air/land battle.[Ref. 7:p.28]

Only three of the TAOs (close air support, interdiction, and airlift) will be looked at in the application of subjective transfer function approach to evaluating the impact of tactical air C^2 on the tactical air force. Figure 9 shows the tactical air force and the components which make up tactical air operations.

E. ELEMENTS OF TACTICAL AIR C^2

Tactical air C^2 may be viewed as constituted from the elements of doctrine, organizational structure, procedures, personnel, facilities, equipment, and communications. These elements give to those responsible at each level of command the ability to carry out the functions of planning, directing, and controlling necessary to accomplish their purpose of meeting mission objectives through the performance of tactical air operations.[Ref. 6:p.3]

For a commander at any level, the reason a C^2 system exists at all is to support his battle management decision making and command functions. Although, these decisions may vary with time, threat and hierarchial level, the basic decision process has been shown to be relatively invariant. A strong conclusion from this is that tactical air C^2 can be divided into a tactical air C^2 decision process and a tactical air C^2 system.[Ref. 8:p.5] Figure 10 shows the three areas of the tactical air force and the outcome of the employment of tactical air force- major military actions and the breakdown of tactical air C^2 if the decision process is assumed to be relatively invariant.

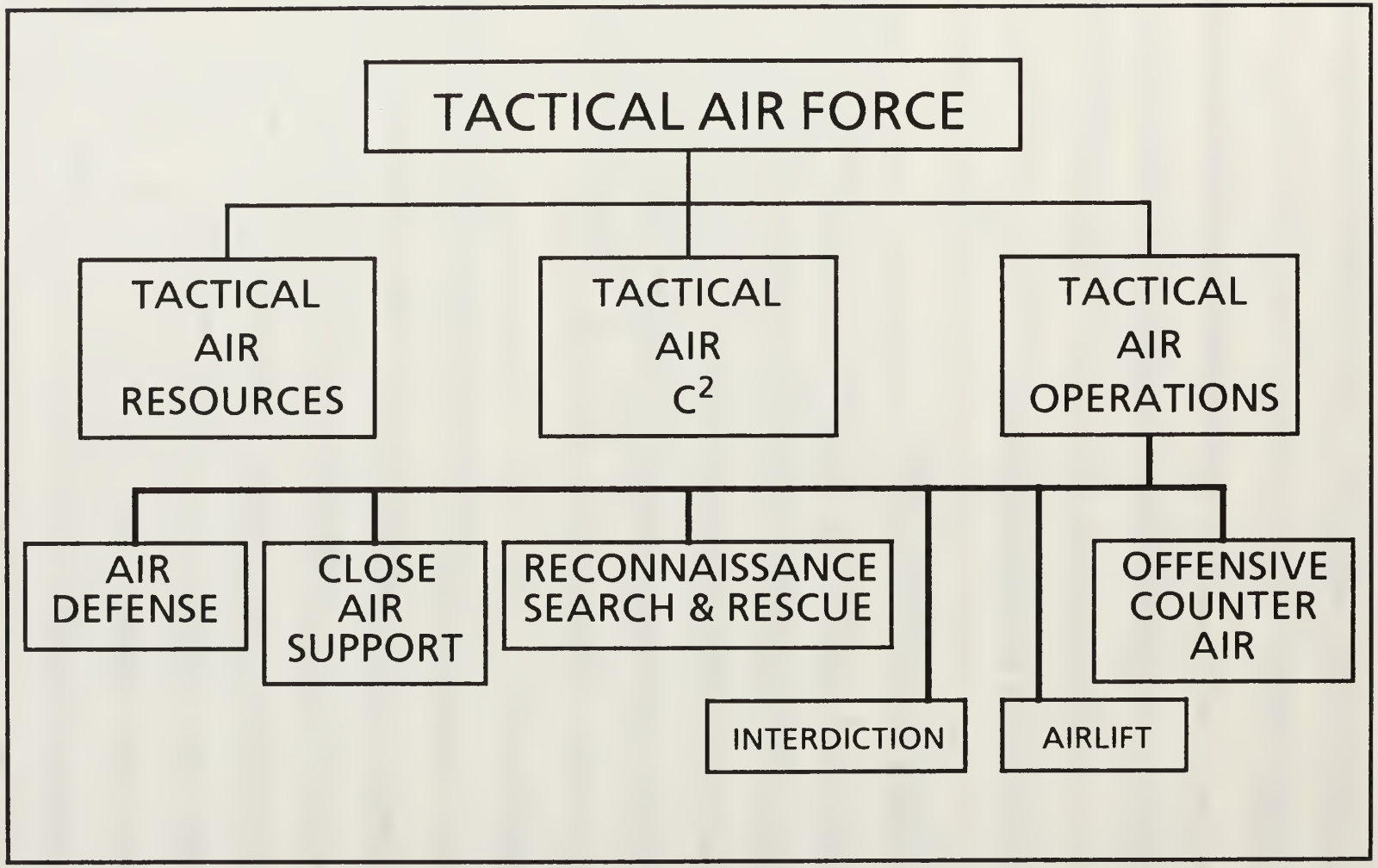

Figure 9. A Tactical Air Force Structure: Tactical Air Operations.

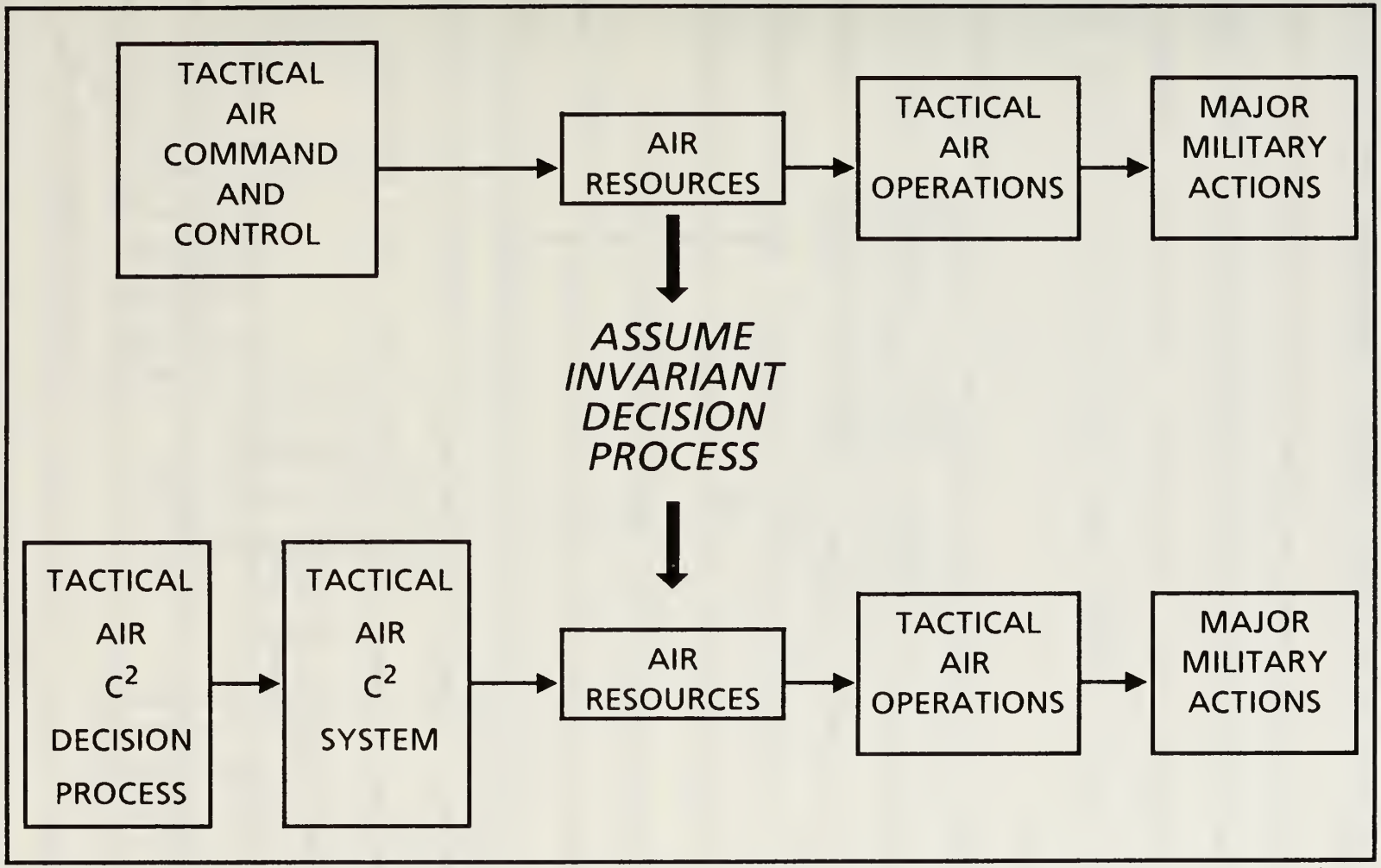

Figure 10. A Tactical Air Force Structure: Assuming an Invariant Decision Making Process.[Ref. 8:pp.4,6]

This thesis concentrates on the tactical air C^2 system aspect of tactical air C^2. Joseph G. Wohl's article "Force Management Decision Requirements for Tactical Command and Control" attempts to structure and define the process of decision making for force management in air force tactical C^2. Wohl's article is referenced to acknowledge the two categories which tactical air C2 can be divided into. [Ref. 9:p.618]

As stated in the previous section, tactical air C^2 provides the linkage between operational requirements and the tactical air resources. In light of this, tactical air C^2 meets tactical air operational requirements with the application of tactical air resources. Tactical air C^2 is also used for operational management of the tactical air resources. Tactical air resources must be maintained at a status capable of supporting the application decisions and the resultant tactical air operations (see Figure 7).[Ref. 6:p.2]

In order to meet mission objectives, it requires the management of effective tactical air operations and the selection of the most appropriate tactical air operations to meet those objectives. An example of this is choosing and conducting tactical air operations in order to have a favorable impact on military actions, such as land battles. This interrelationship is depicted in Figure 11.

The TACS, as previously mentioned, manages the employment of the tactical air forces. It determines which enemy targets to destroy, which information to collect, where and what to airlift, and directs specific wings to perform specific tasks at specific times. TACS includes a network of operation centers, communication systems, and ground and airborne radars. It maintains, as complete as possible, a picture of the unfolding air and

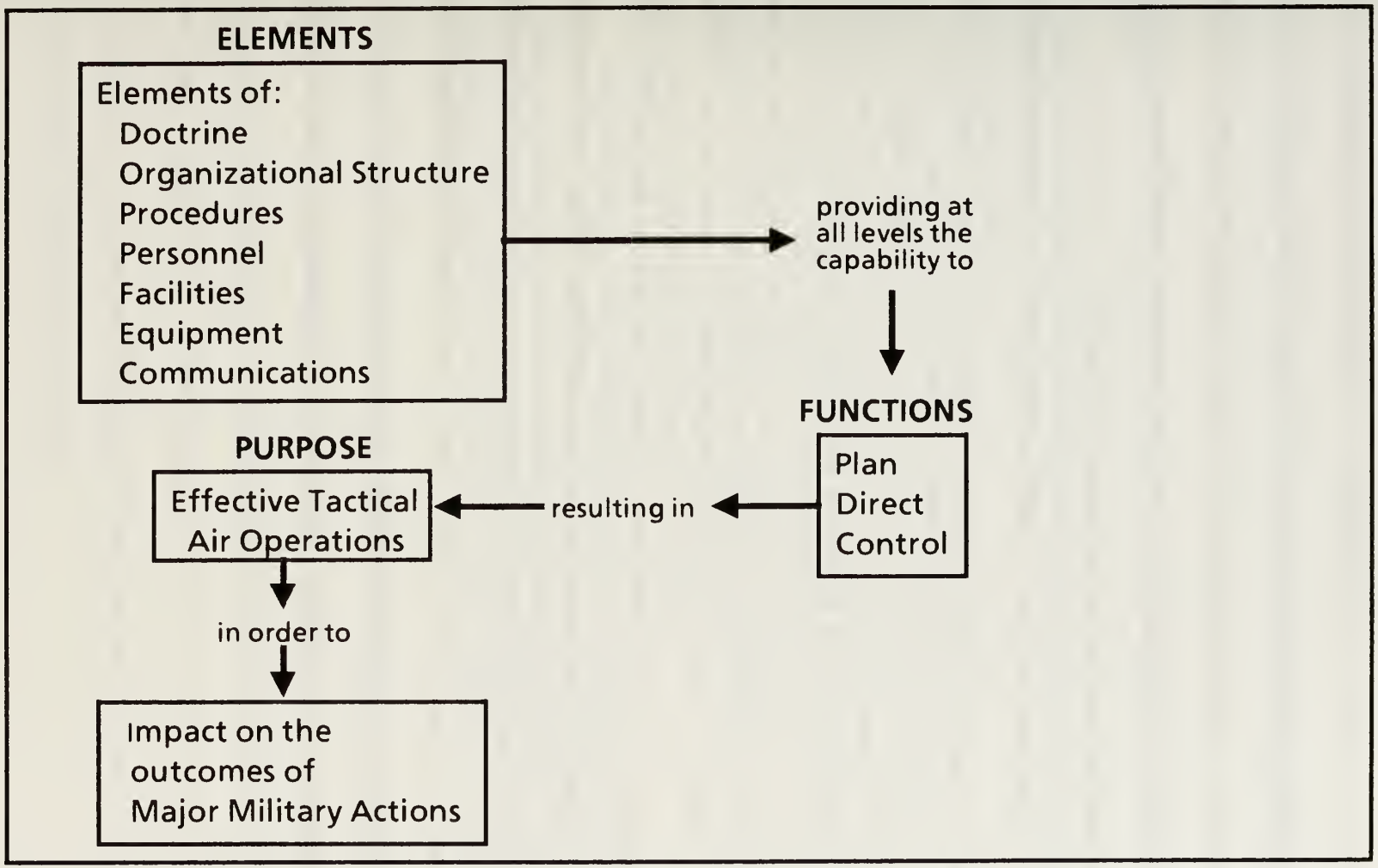

Figure 11. Tactical Air Command and Control.[Ref. 6:p.4]

land battles. Also, TACS shows the posture of unengaged forces, both friendly and enemy, by processing the friendly and enemy information provided to it. Senior officers in the TACS make the force employment decisions and direct the wings accordingly, by taking into consideration national and military plans and objectives and the above picture, provided by TACS, of the air and land battles.[Ref. 7:p.27]

Force employment decisions are made in two contexts: 1) a future operational time period and 2) a present operational time period. The decision of how to employ the force in a future operational time period (usually, the next day) is called *planning*. In each period, the employment of the tactical air force, which is expected to be available, is planned for the following period. While the plan is being executed, there are decisions which are required to adjust the planned employment in response to the currently perceived situations that differ from those projected at the time the plan was made. This process of employment decision making is called *controlling*. Notice in both cases, the decisions specify operational missions to be flown by the tactical aircraft and the wings are directed to perform these missions. This process of executing the planning and controlling functions is called *directing*. Thus, tactical air C^2 can be viewed as performing three main functions: 1) planning, 2) directing, and 3) controlling.[Ref. 7:p.27]

F. MOFEs FOR THE TACTICAL AIR FORCE

As stated previously, tactical air forces affect the progress and outcome of military events by flying or having the potential to fly combat missions. These missions are categorized into TAOs, which indicate primary mission objectives. From this,

> ...tactical Air Force employment in general can be thought of as the performance of tactical air operations, and the effectiveness of force employment can be thought of as the effectiveness of appropriate TAOs in affecting the course of military events.[Ref. 7:p.27]

In his article "The Measurement of Combat Effectiveness", Philip Hayward proposes that "...a satisfactory quantitative measure of the combat effectiveness of a military force is its probability of success in combat." [Ref. 10:p.618] Hayward goes further and stresses the importance of defining the relationship of the probability of success with force capability, environment, and mission.

> The probability of success depends not only on the capabilities of the specified force but also on the nature of the enemy, the combat environment, and the mission. Since it is impractical to measure combat effectiveness experimentally, i.e., in actual combat, military judgement must be called on to specify the relation between the probability of success and the parameters of force capability, environment, and mission.[Ref. 10:p.314]

Simply stated, the measure of force effectiveness (MOFE) for a tactical air force can be viewed as the probability of success of its TAOs in combat. When measuring the effectiveness of an organization, it is accomplished through an analysis of the data on its performance under actual operating conditions. However, in this situation, this is not a viable option, at least not yet. Realizing live data from an actual combat scenario is not a realistic check, Hayward goes on to state

> ...the validity of the results...can be checked in only one way-by actual combat. In the absence of such a check, the most that can be claimed for any proposed measure of combat effectiveness is not that it is 'correct', but that the arguments on which it is based are clear (i.e., capable of being analyzed and debated in a meaningful way), logically consistent, and in general accord with the judgements of military experts.[Ref. 10:p.322]

It is this approach which will be taken. The TAOs which make up the tactical air force have already been defined. The combined success of each of the these TAOs based on the judgements of military experts is the measure of force effectiveness for the tactical air force, which will be focused on. In particular, the relationship of probability of success with mission will be looked at in detail, while the other areas of force capability and environment will be held constant. However, it is important not to forget that when they are changed, so does the combat effectiveness of a military force (i.e. tactical air force).

G. MOEs FOR TACTICAL AIR C^2

The evaluation of tactical air C^2 can be thought of as the evaluation of the three functions, which comprise tactical air C^2-planning, directing, and controlling. Referring back to Figure 11, the following observations can be made about evaluating tactical air C^2. First, the inputs to the C^2 process are the elements which make up the C^2 system and bound its capability. These inputs have well-defined measurable attributes (e.g., quantity, performance factors, physical characteristics) which together describe a C^2 system. To produce variations in the capabilities of C^2, only the elements can be added to, reconfigured, and modified.[Ref. 6:p.3]

Second, the outputs of the overall C^2 process are effective tactical air operations and favorable effects on major military actions. Thus, the decisive products of C^2 are its contributions to the above outputs.[Ref. 6:p.3]

Finally, the inputs to the C^2 process do not affect the outputs directly, but indirectly through the functions of planning, directing, and controlling, which are human dominated processes.[Ref. 6:p.4]

> Planning, directing, and controlling use the elements and what they provide, but plans, directions, and control actions are the results of decisions made by people.[Ref. 6:p.4]

As will be shown in Chapter IV, the dominant human element in the C^2 process will be important in the methodology used for relating the effectiveness of tactical air C^2 to the force effectiveness of a tactical air force.

In choosing a measure of effectiveness for tactical air C^2, it is desirable that it be in terms of how it affects military events in wartime. Tactical air C^2 affects military events through its impact on the performance of tactical air operations. Finally, the effectiveness of the TAOs is based on how well the three functions of C^2 can be executed.

> Evaluation of the effectiveness of Tactical Air Command and Control...must be in terms of how it can affect the course of military events in wartime...Command and control affects military events only through its effects on the performance of tactical air operations. Command and control...[reflects] the effectiveness of the TAOs... [which] depends in large part on how well they can be planned, directed, and controlled.[Ref. 7:p.28]

Thus, the measure of effectiveness for tactical air C^2 is chosen to be the composite measure of the success of each of the three functions: planning, directing, and controlling.

IV. A METHODOLOGY FOR RELATING MOEs TO MOFEs

A. OVERVIEW

In Chapter II, a hierarchy of MOEVs was presented. Each MOEV was defined and a relationship to the military system and the hierarchy of elements, comprising it was established. In Chapter III, a "complex system", tactical air forces was discussed in detail. The major elements comprising tactical air forces were described. In particular, the element of tactical air C^2 was broken down into it's component elements. A measure of force effectiveness (MOFE) was derived for tactical air forces and a measure of effectiveness (MOE) for tactical air C^2. In Chapter IV, a methodology for relating MOEVs to each other is described. An overview of the STF approach will be followed by a more detailed discussion of each of the five steps, which make up the STF approach.

B. THE SUBJECTIVE TRANSFER FUNCTION (STF) APPROACH

A complex system is composed of a number of different components. Examples of complex systems include military C^2, education, transportation, management, and tactical air forces. Before trying to understand what affects the outcomes of a complex system as a whole, the cause and effect relationships between the components of the system need to be understood. Once they are known and the system outcomes are identified then the effectiveness of the system can be assessed.[Ref. 7:p.1]

A system's effectiveness can be evaluated using both qualitative and quantitative techniques. In the area of qualitative techniques, a system's effectiveness is often evaluated using subjective measurement techniques. These techniques, typically, ask "experts" to respond to questions about particular aspects of a system using a numerical scale. Because these measurement techniques require the interpretation of expert responses in terms of processes that cannot be observed directly, it is called "subjective". Response interpretations will usually concern the following: 1) the subjective scale values associated with specified characteristics of the system and/or 2) how subjective values of the characteristics of the system affect the judgements of a system's outcome (this requires the expert's underlying judgement model to be specified). Usually, response interpretations are used as input to operational and management decisions.[Ref. 7:p.1]

One subjective measurement technique is the STF approach, which relies on "expert" judgement for analyzing complex systems where many factors directly or indirectly affect the outcomes of the system. It is desirable to define a system in terms of its factors and outcomes. A major problem in any evaluation procedure is to accomplish this in such a way that causal hypotheses about the effects of those factors on the outcomes can be tested and rejected if the data do not support them. However, hypothesis testing is an integral part of the STF approach, which is based on the algebraic modeling approach to measurement. The STF approach relies on algebraic models imbedded into each experimental design to model the subjective values which the experts attach to the information they are judging. The scale values are a result of the theory (algebraic models) and have meaning if and only if the judgement theory passes stringent

tests of its predictions. In other words the algebraic model designed into the experiment is verified only if it correctly predicts the expert responses (subjective values).[Ref. 11:p.1]

As stated above, in the STF approach complex systems are analyzed from the perspective of the "expert". By definition, the expert knows and understands the system. Usually, different groups of experts know about different aspects of the system. From each group, experts make judgements about outcomes resulting from the factors (which they are knowledgeable on). These judgements are based on what they would expect under different descriptions of system capabilities. For each expert group, the judgement theory (STF) specifies the effects of different system capabilities on the judged outcomes. It is the set of STFs across the expert groups, which links the outcomes associated with different tasks within the system to an outcome(s) corresponding to a measure of the overall system effectiveness.[Ref. 11:p.1]

The following steps outline the STF approach:

1. Define the complex system representation.
2. Conduct judgement experiments.
3 Determine the STFs and the final structure.
4. Evaluate the system capabilities using the model.[Ref. 11:p.1]

Thus, the STF approach is based primarily on hypothesis testing, which uses algebraic modeling to predict the responses of a group of experts on the relationship of a group of factors, whose causal relationship produces an outcome. Each judgement experiment is designed to test a hypothesis. The hypotheses are tested on effects of the defined system components (factors) on specified outcomes. It is important to understand

that the STFs incorporated into each experiment design are hypothesized models, which the researcher hopes will specify the nature of these effects. After empirical support has been collected for the effects of all the hypothesized components on the judged outcomes, the result is a final representation for the complex system. Judgement experiments are continued for each experimental unit (group of factors and outcome) until an STF is found, which predicts the expert responses (which represent the relationship between the components (factors)). The STFs are considered "appropriate" only when their predictions are verified by the data (expert responses). The next section outlines the first step in the STF approach-defining a complex system representation.[Ref. 6:p.24]

1. Defining a Complex System Representation

A complex system structure shows the basic elements, called *factors*, which make up a system. It also shows the direct and indirect effects of the factors on the outcomes of the system. The causal links between the factors and the outcomes of the system are determined by the STFs. Since, hypotheses can only be tested when experiments have been appropriately designed, alternative structures and STFs must be hypothesized before the data is collected. Developing an initial structure of a complex system requires identifying the system effectiveness outcomes of interest, and postulating the factors thought to affect them. This section explains the procedures for developing hypotheses of system structures and STFs, which together form a representation of the complex system being evaluated.[Ref. 11:p.3]

a. Structural Hypotheses

The development of structural hypotheses requires the input of system experts. As different people are expert in different parts of the system, a structure is developed for each part of the system in conjunction with the people who are expert in that part of the system. Each group of experts helps identify and define important and realistic system outcomes and factors, which can be manipulated in experimental designs.[Ref. 11:p.3]

(1) Identifying System Outcomes and Factors. First, system outcomes must be identified, which provide the important external measures of the system's effectiveness. Next, the system factors thought to directly affect these outcomes are identified. Some or all of these factors may represent outcomes produced within the system. These are called *suboutcomes* and are affected by other system factors. The system factors are identified for suboutcomes until all suboutcomes are affected only by factors representing system input characteristics. When this is accomplished a hierarchial causal representation of the system is formed. The system input characteristics or basic system features are called *primitive factors*.[Ref. 11:p.3]

For an example, let's look at a structure for a tactical C^2 process. As depicted in Figure 12, the structure contains one factor/outcome set in the top part of the figure and another factor/outcome set in the bottom part of the figure. These sets are called *experimental units*. These two experimental units correspond to two different groups of experts. The group of experts corresponding to experimental unit 1 (top part of Figure 12) performs the immediate targeting task of pairing tactical aircraft with

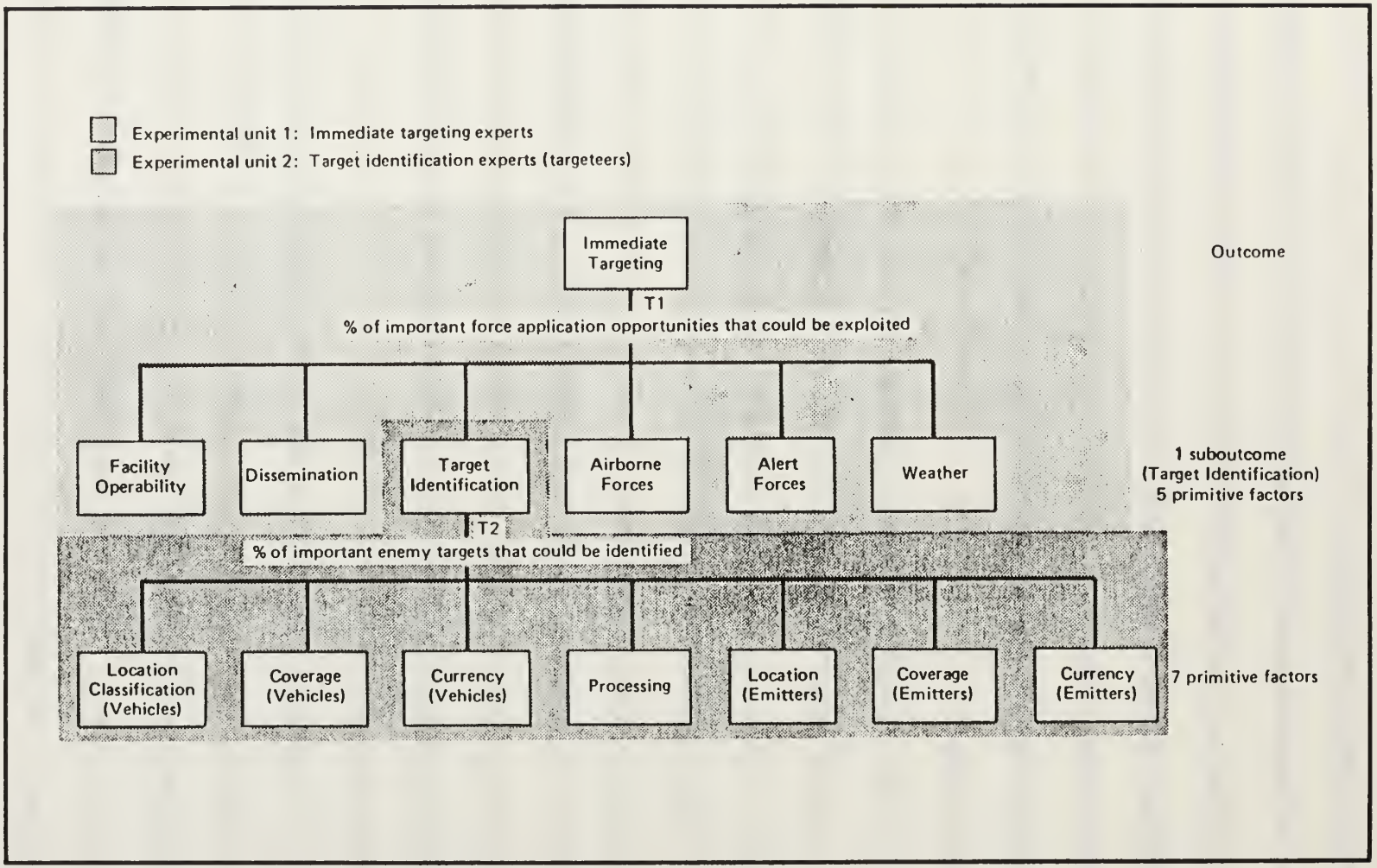

Figure 12. Hypothesized Immediate Targeting Structure.[Ref. 11:p.4]

important enemy ground force targets in a timely manner. In this example, the overall measure of system effectiveness is how well U. S. Air Force officers perceive they can perform their immediate targeting task under various conditions relating to the information and equipment they work with. The group of experts corresponding to experimental unit 2 (bottom part of Figure 12) perform the identifying enemy targets task. The factors hypothesized by each group of experts to affect an outcome or suboutcome must not exceed the span of knowledge of the group. Notice in experimental unit 1, there are six factors hypothesized to affect Immediate Targeting directly. In experimental unit 2, there are seven factors hypothesized to affect Immediate Targeting indirectly through the suboutcome of Target Identification. The definitions of system factors and outcomes for this immediate targeting structure is shown in TABLE I.[Ref. 11:pp.4-5]

It is necessary to identify the factor levels for each factor, since the factors will be manipulated in the experimental designs. For each of the factors shown in Figure 12, factor levels were selected and are shown in TABLE I. Factor levels should cover the range from the worst to the best capability which could be expected over the time period of interest. This is an important feature if future conditions or characteristics of future systems are going to be included in the model for the purpose of evaluation. The following areas may guide the selections of factor levels between the worst and best capabilities: descriptions of equipment being considered for research, production, or purchase, and descriptions of existing equipment capabilities. Usually, three to five levels for a factor are sufficient for experimental purposes.[Ref. 11:p.5]

TABLE I. DEFINITIONS OF FACTORS AND OUTCOMES FOR IMMEDIATE TARGETING STRUCTURE SHOWN IN FIGURES 12 AND 13.[Ref. 11:p5]

A. Experimental Unit 1 (Immediate Targeting Experts)

Judged Outcome: The percent of force application opportunities that could be exploited in a timely manner

Factor Definitions	Factor Levels
Target Identification (percent of important force elements identified)	90 60 30 10
Facility Operability (percent of immediate targeting activities that can be supported by the facility)	90 60 30 10
Alert Forces (status of the Alert Forces accessible in the C^2 facility)	90 60 30 10
Airborne Forces (status of the airborne forces accessible in the C^2 facility)	90 60 30 10
Weather (currency of the reliable weather information)	15 min., 1 hr., 3 hrs., 12 hrs.
Dissemination (percent of the forces that can be tasked in a timely manner)	90 60 30 10

TABLE I. (Continued) DEFINITIONS OF FACTORS AND OUTCOMES FOR IMMEDIATE TARGETING STRUCTURE SHOWN IN FIGURES 12 AND 13.[Ref. 11:p6]

B. Experimental Unit 2 (Targeting Experts)

Judged Outcome: The percent of important enemy targets that could be identified in a timely manner

Factor Definitions	Factor Levels
Vehicle Location/Classification (ability of sensor systems to locate and classify enemy vehicles)	Locate and classify in all weather Locate (not classify) in all weather Locate and classify in clear weather
Vehicle Coverage (percent of enemy vehicles that have been observed)	90 60 30 10
Vehicle Currency (time interval between the observation of enemy vehicles and the data's availability for processing)	5 min., 15 min., 30 min., 1 hr.
Processing (the means by which enemy vehicle and emitter information is interpreted)	Fully computerized interpretation. Human uses computer to graphically display information; human interpretation. Human uses computer to sort textual information; human interpretation. Human sorts hard copy, textual information; human interpretation.
Emitter Location Accuracy (accuracy with which enemy emitters are located)	10m, 100m, 1000m
Emitter Coverage (percent of the enemy emitters that have been observed)	90 60 30 10
Emitter Currency (time interval between the observation of emitters and the data's availability for processing)	5 min., 15 min., 30 min., 1 hr.

The structural hypothesis (Figure 12) implies immediate targeting depends on the ability of the targeteers to identify important enemy targets. A contributing factor in experimental unit 1 and outcome in experimental unit 2 is Target Identification. In cases like this, where a factor serves a dual purpose, it is necessary to define the factor in the same terms for each group of experts to satisfy the transfer function of the STF approach.[Ref. 11:p.7]

Two STFs are hypothesized in the structural representation shown in Figure 12. The first STF, T1, specifies the causal link for the six factors affecting the immediate targeting outcome. The second STF, T2, specifies the causal link among the seven factors affecting Target Identification. Because theses functions are models of judgement processes, which cannot be directly observed occurring between the time a stimulus is perceived and the time a response occurs, they are called *subjective.* In the STF approach a *stimulus* is a description of the system's capabilities and a *response* is a judgement of a task performance. These subjective functions are called *transfer functions* because after their function forms are determined and are being computed to evaluate a particular system, the output of one function is transferred for use as an input value to the function above it. In this example, the output of T2 in Figure 12 identifies the target identification factor level needed to determine that factor's subjective input value to T1.[Ref. 11:p.7]

(2) Identifying Alternative Structures. Alternative structures are alternative hypotheses about the number of STFs linking the system factors to the system outcomes. Alternative structural hypotheses are formed from different hypotheses about

how the group of experts combines the information included in the description of a system's capabilities. Hypotheses may result from interaction with the group of experts during the development of the structure. Finally, hypotheses may come when data is analyzed.[Ref. 11:p.10]

Referring back to the previous example of the immediate targeting structure (Figure 12), an alternative structural hypothesis for experimental unit 2 is illustrated in the bottom part of Figure 13. This alternative structure suggests that Air Force targeteers combine information about enemy vehicles. The subjective values of those outputs are taken and combined with their value, which is associated with the processing capability factor. Three STFS (T3, T4, and T5) are required for this alternative structural hypothesis. By inserting into the structure two intermediary factors: Vehicles and Emitters, two separate combination processes are represented. *Intermediary factors* are factors not identified by factor levels because they are not manipulated in the experimental designs. They represent the process of separately combining a subset of factors in the experimental unit.[Ref. 11:pp.11-12]

If we choose other structural hypotheses, the number of paths and STFs will change. An example of an alternative structure for experimental unit 1 could result from a hypothesis in which an intermediate targeting expert combines information about the three factors concerning their friendly forces (Alert Forces, Airborne Forces, and Weather) separately from information about their other capabilities of Target Identification, Facility Operability, and Dissemination). By inserting an intermediary

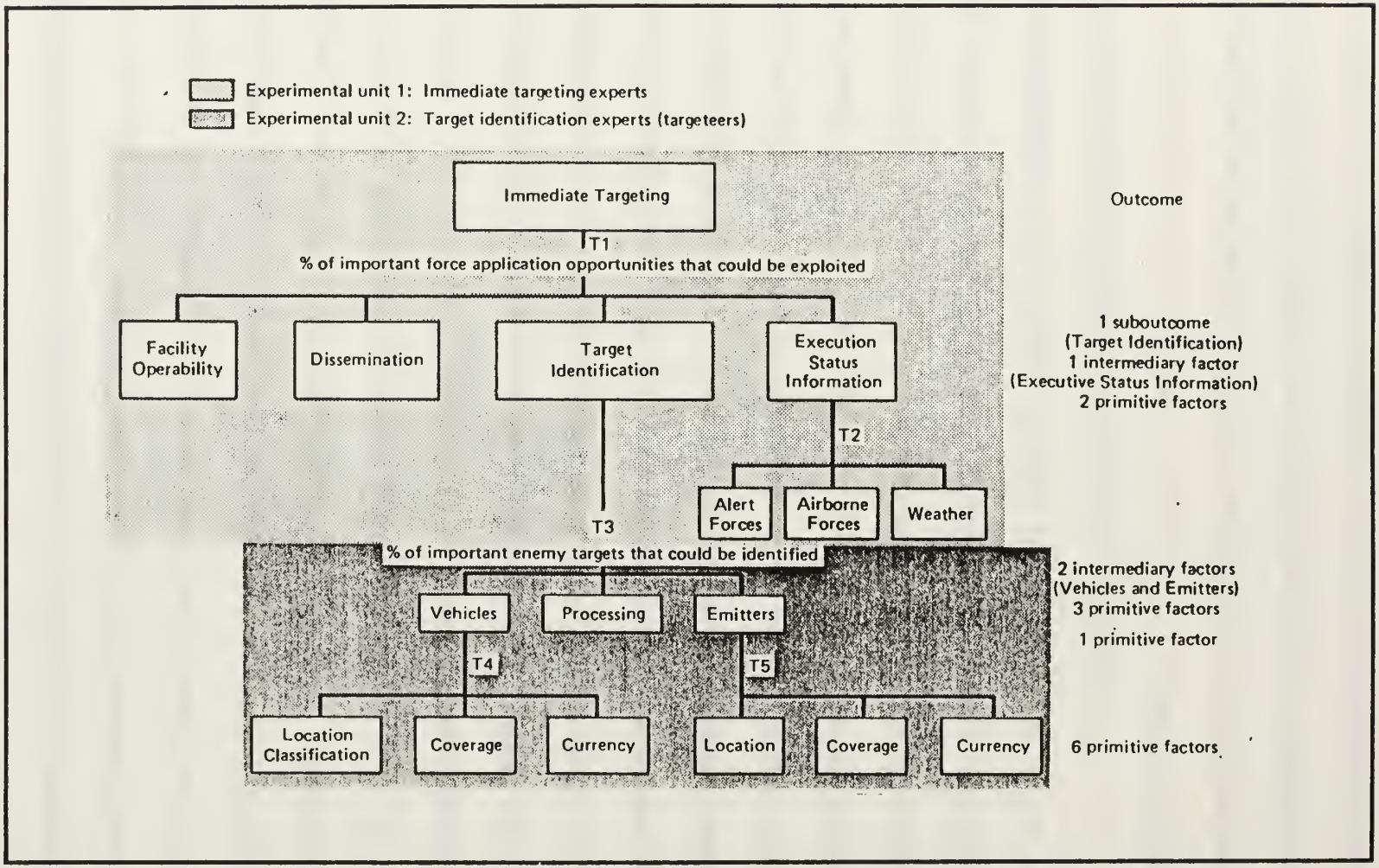

Figure 13. Alternative Structure to Figure 12. (Shows Three Intermediary Factors.)[Ref. 11:p.11]

factor called Execution Status Information in this part of the structure we can show this. Two STFs (T1 and T2) for experimental unit 1 are required for this alternative structure. Again, if we choose another structural hypothesis the number of paths and STFs will change.[Ref. 11:p.12]

Specification of as many appropriate alternative combination hypotheses before data collection is important to ensure the experimental designs provide a basis for knowing which of the alternative structures accounts for the data the best. This is especially important in larger experimental units which contain four to six factors.[Ref. 11:p.12]

The above example is fairly small in size. The immediate targeting structure in Figure 12 has only 13 factors to be manipulated in experimental designs. It has only one suboutcome and only one outcome. A lot of systems are larger. One example, the tactical air C^2 system depicted in Figure 14, is composed of 12 STFs, 25 factors, six intermediary factors, six outcomes, and a final outcome. This only shows one-third of the structure hypothesized to affect the final outcome of the land battle.[Ref. 11:p.12]

b. STF Hypotheses

A STF hypothesis is an algebraic model. It specifies the subjective (i.e. unobserved) processes between the perception of a stimulus, such as a questionnaire item and the occurrence of a response. Figure 15 shows an outline of these processes involved in forming an STF hypothesis. Although this outline shows three factors, it could be

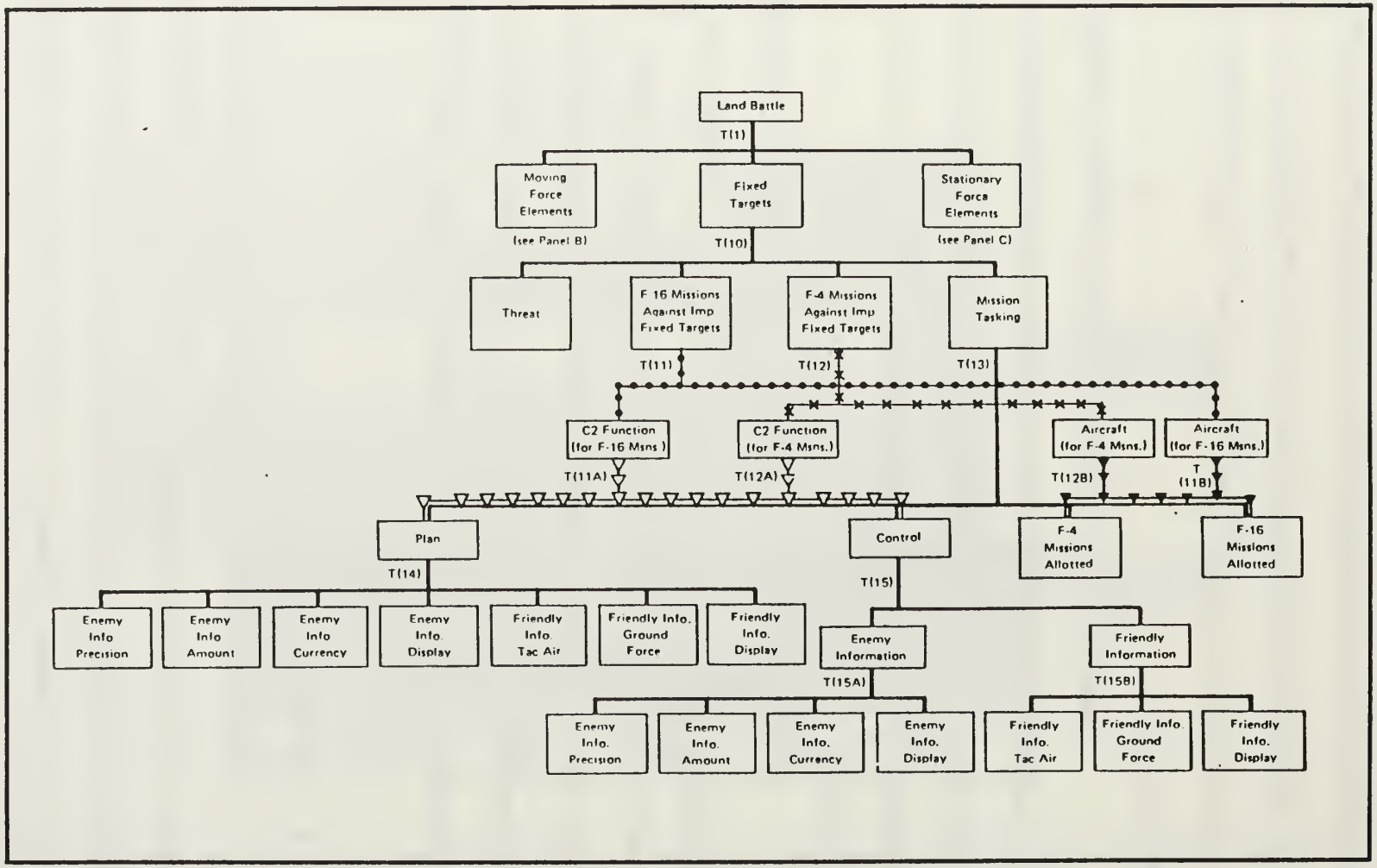

Figure 14. Final Structure of a C² System.[Ref. 11:p.14]

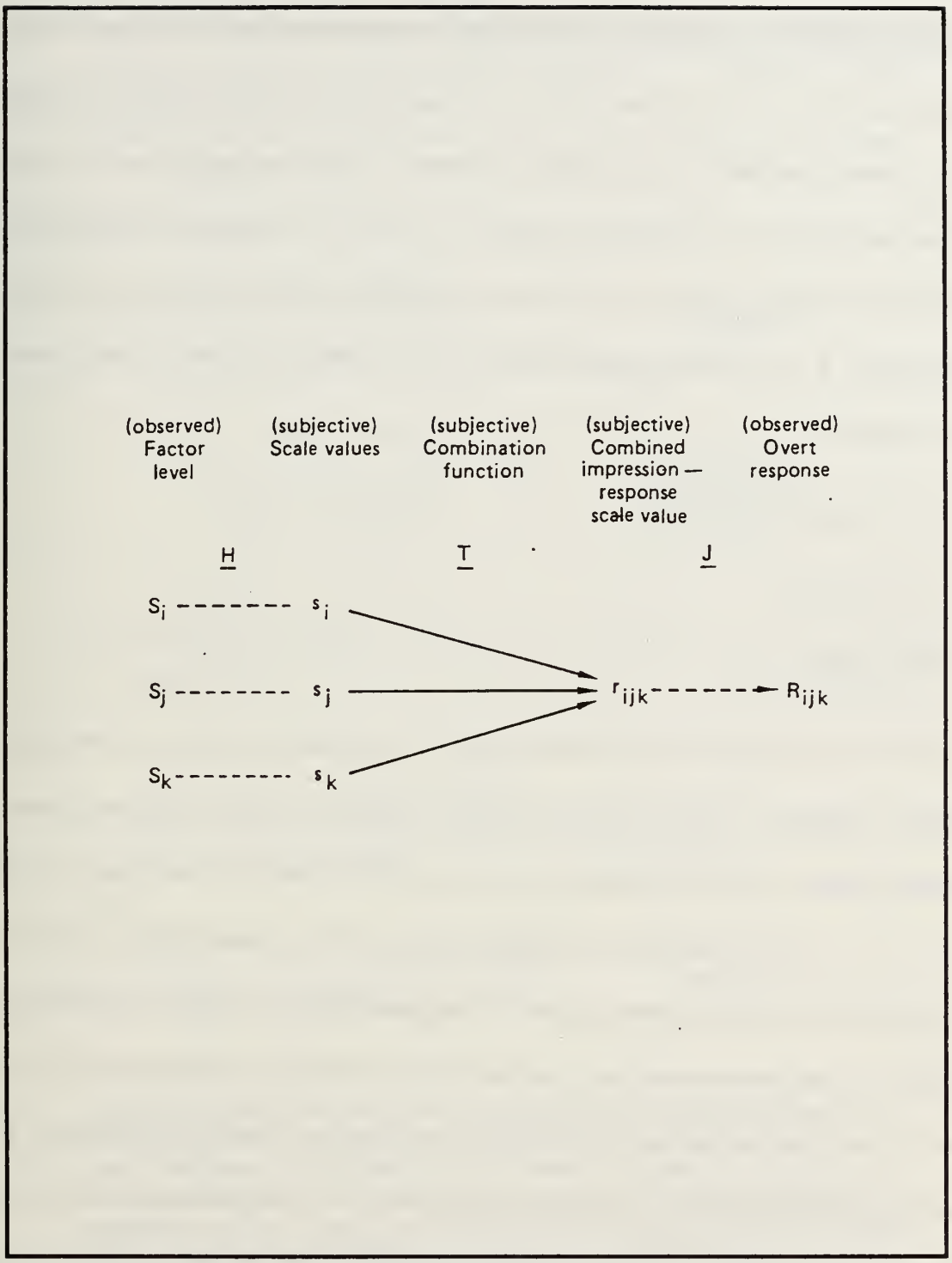

Figure 15. Outline of Subjective Measurement.[Ref. 11:p.15]

expanded to include an infinite number. On the left, the observed stimuli would be factor levels from three different factors. The outline shows the expert first transforms each factor level (S_i, S_j, S_k) to a scale value (s_i, s_j, s_k) using some function, **H**, which is referred to as a *utility function* or *psychophysical function.* Next, the expert combines these values according to a combination function, **T**, to form an integrated impression, $\mathbf{r_{ijk}}$. Finally, the expert transforms the psychophysical impression to an overt response, $\mathbf{R_{ijk}}$ by the function, **J**. The above judgement stages can be written in the form of the following three equations:

$$\mathbf{s_i = H(S_i)} \text{ for stimulus i,} \quad \textbf{(1)}$$

$$\mathbf{r_{ijk} = T(s_i,s_j,s_k)} \quad \textbf{(2)}$$

$$\mathbf{R_{ijk} = J(r_{ijk})} \quad \textbf{(3)}$$

An STF (**T** in Figure 15 and Equation (2)) is a perceptual theory of the expert's judgement process. The STF shows how the subjective values the expert places on different factors affect his judgements, such as ability to perform a task.[Ref. 11:p.15]

Some algebraic functions which could be considered as STFs at the beginning of a complex system analysis are listed in TABLE II. Judgement literature and previous research in the area being analyzed help in determining which functions to start with. Note that the functions in TABLE II are based on three factors, however, they could easily be rewritten to include any number of factors. Equations (1)-(3) define the **r**, **w**, and **s** parameters. The initial estimate parameters, $\mathbf{w_0}$ and $\mathbf{s_0}$, are what the response

TABLE II. POSSIBLE STFs.[Ref. 11:p.16]

A. Noninteractive functions	
$r = s_{A_j} + s_{B_j} + s_{C_k} + a$	Additive
$r = \dfrac{w_A s_{A_j} + w_B s_{B_j} + w_C s_{C_k}}{w_A + w_B + w_C}$	Averaging
$r = \dfrac{w_0 s_0 + w_A s_{A_j} + w_B s_{B_j} + w_C s_{C_k}}{w_0 + w_A + w_B + w_C}$	Relative-weight (averaging with initial impression)
B. Interactive functions	
$r = s_{A_i} s_{B_j} s_{C_k} + a$	Multiplicative
$r = \dfrac{w_0 s_0 + w_A s_{A_i} + w_B s_{B_j} + w_C s_{C_k}}{w_0 + w_A + w_B + w_C} + \omega(s_{MAX} - s_{MIN})$	Range
$r = \dfrac{w_0 s_0 + w_{A_i} s_{A_i} + w_{B_j} s_{B_j} + w_{C_k} s_{C_k}}{w_0 + w_{A_i} + w_{B_j} + w_{C_k}}$	Differential-weight

*All equations are for three factors: A, B, and C. The parameters of the functions are as follows: s_{A_i}, s_{B_j}, and s_{C_k} are the subjective values for the i^{th}, j^{th}, and k^{th} levels of factors A, B, and C, respectively; w_A, w_B, and w_C, are the weights associated with factors A, B, and C, respectively (a subscript is added when the scale value varies with the factor level); r is the subjective response; w_0 and s_0 are the weight and scale value associated with the initial impression (what the response would be in the absence of specific information); ω denotes the weight of the range term; and a is an additive constant.

would be if specific information was unavailable. The **J** function in Equation (3) relates subjective responses, **r**, to observed responses. Its determination is discussed later on in the section on scale-free design.[Ref. 11:p.15]

Each of these functions (TABLE II) makes a different prediction about the pattern the judgement data should follow when appropriate experimental designs are used. For example, the functions in the upper panel of TABLE II all predict no interactions between the factors. The functions in the lower panel all predict interactions between the factors. The functions within each panel make other predictions with respect to the judgement data.[Ref. 11:p.15]

So, the experimental designs allow for adequate tests among their predictions. The forms of the STFs considered as possible explanations of the effects of factors on judged outcomes must be specified in advance. Subjective values associated with the factors and outcomes (the **r**, **w**, and **s** parameters for the functions in TABLE II) are known when the tests of an STF support it as an appropriate explanation of the experts' judgements and are the least-squares estimates of parameters of the function. Because no assumptions are needed on the distribution of the responses, least-squares estimates are usually preferred to maximum likelihood procedures for estimating parameters.[Ref 11:pp.15-16]

2. Conducting Judgement Experiments

The second step in the STF approach is to carefully select experimental designs to allow tests between the hypotheses. These experimental designs are translated into a questionnaire, which is fielded to the appropriate expert respondents.[Ref. 11:p.17]

a. Designing Experiments to Test Hypotheses

The experimental design is crucial to ensuring testing is accomplished among the unique predictions of the STFs being considered. For the most part, the experimental design is guided by the knowledge of the researcher on the predictions of the STFs. The researcher is guided by the alternative hypotheses in selecting combinations of factors, when the experimental units have four or more factors.[Ref. 11:p.17]

Experimental designs produce factor level combinations, which are descriptions of the system's capabilities. The following is an example of a description for experimental unit 1 (Figure 12):

> 30 percent of the important 2nd echelon force elements are identified in a timely fashion. Facilities can support 60 percent of the necessary immediate targeting activities. Tasking can be correctly communicated to 60 percent of the forces in time. There is timely access to the status of 10 percent of the Alert and Airborne forces. Weather data are three hours old.[Ref. 11:p.17]

This represents a possible questionnaire item for experimental unit 1. Other items might ask experts to determine how well they could perform their immediate targeting task in a C^2 system which has these capabilities.[Ref. 11:p.17]

(1) Factorial Designs: Tests Between Interactive and Noninteractive Functions. Fully crossed factorial designs are the backbone of other designs. They depict predictions which can be assessed from factorial designs. Every level of every factor, in a fully crossed factorial design, is combined with every level of every other factor in the design. For example, a two-way factorial design of the Alert and Airborne Forces factors (Figure 12 and TABLE I) is shown in Figure 16A. Since there are two factors, each with

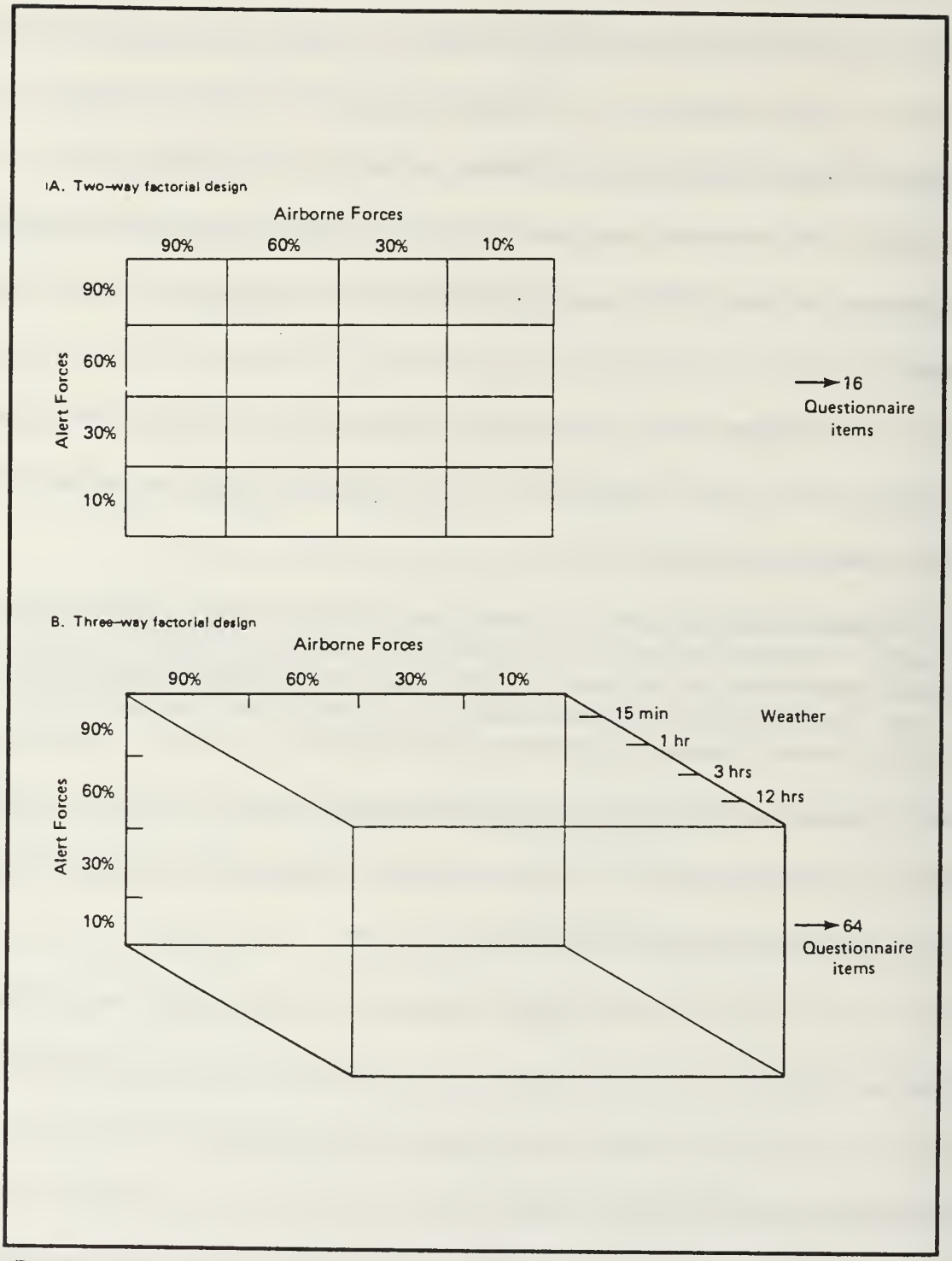

Figure 16. Example of a Two-Way (**A**) and a Three-Way (**B**) Factorial Design.[Ref. 11:p.18]

four factor levels, a 4 x 4 design is produced. There are 16 cells in the design, each of which represents a situation described in a questionnaire item.[Ref. 11:p.17]

Figure 16B shows a complete three-way factorial design of the alert forces, airborne forces, and weather factors shown in Figure 12 and described in TABLE I. Since each of these factors has four factor levels, the design is 4 x 4 x 4 with 64 cells or questionnaire items. As the number of factors and/or factor levels increase, the number of questionnaire items generated from the fully crossed design increase rapidly.[Ref. 11:p.17]

Main effects and interaction effects among the factors can be assessed by factorial designs. After repeated tests, if a proposed factor has no effect on judgements, its appropriate parameter is set to zero in the STF. Its appropriate parameter is either its weight or the scale value as indicated by the data and STF. A basis for choosing between noninteractive and interactive functions is provided by tests of interaction effects for those factors involved in the test. This is illustrated in Figure 17, which shows hypothetical data obtained from 2-way factorial designs (Figures 17A and 17B) and for data obtained from 2 x 2 designs (Figures 17C and 17D). These two examples assume the function, **J** (Equation (3)) is linear. This assumption is discussed further in the section on scale-free tests.[Ref. 11:pp.18-19]

The data shown in either Figure 17A or Figure 17B could have been gathered from the 4 x 4 design (Figure 16A). For diagnostic purposes the data is graphically displayed in Figure 17. The mean response is plotted as a function of the Air Forces factor level with a separate curve for each level of the Alert Forces factor. An

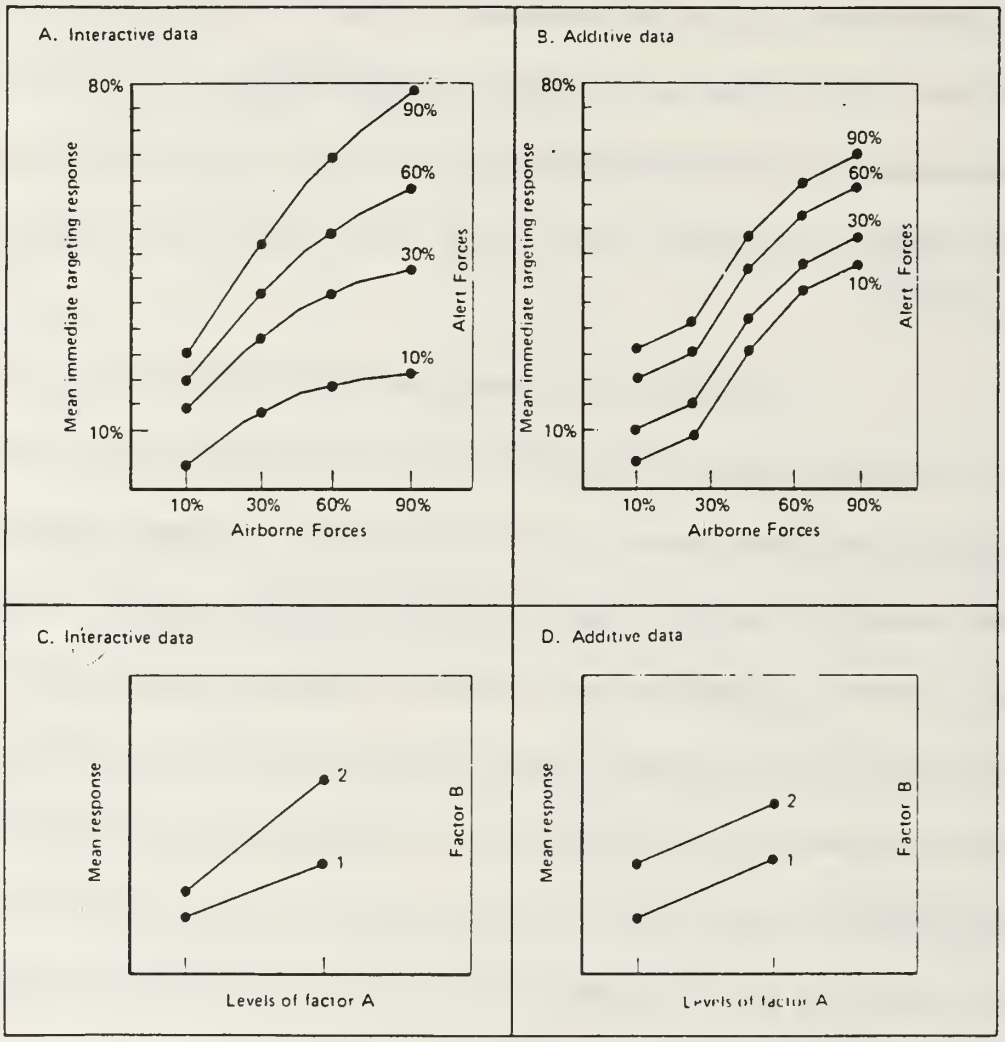

Figure 17. Possible Hypothetical Data Obtained From Two-Way Factorial Designs. (Panels A and B Correspond to Data Obtained From 2 x 2 designs.)[Ref. 11:p.20]

analysis of variance would be significant, since the data shows a large interaction. However, this test would not provide any information on the nature of the interaction. This is critical for deciding if the following questions should be answered: 1) should the interaction be interpreted, 2) if so, how should it be interpreted, and 3) which function(s) account for the data. Information about these questions is provided in Figure 17A. It shows that the interaction is systematic and not the result of a lot of scatter which could cause points on different curves to cross or result in outliers. Also, it shows the interaction is divergent, which means the expert respondents are saying that the less information about the status of the Airborne Forces, the less of a difference it makes how much information is available about the status of the Alert Forces and vice versa. The interactive functions, range and multiplicative (TABLE II), account for data which has this pattern. Through additional experimental designs and graphic diagnostics, it is possible to choose between the two interactive functions.[Ref. 11:p.19]

As more factors and factor levels are included in the design, tests among proposed interactive and noninteractive functions become more powerful. This is illustrated by comparing the two lower panels (Figures 17C and 17D) with the two upper panels (Figures 17A and 17B). The hypothetical data shown in the two lower panels are for two levels of each factor. In larger designs (the two upper panels), there are more ways for curves to show nonparallelism. The idea that one of the class of additive functions is the appropriate combination process (STF) is more acceptable if parallelism is observed. When systematic instructions are observed with larger designs, it is more acceptable for the observed systematic trends to be interpreted.[Ref. 11:p.19]

A basis for the discernment between interactive and noninteractive effects is provided by factorial designs. But, they are not adequate for differentiating between the models within either class.[Ref. 11:p.19]

(2) Factorial Designs: Extensions Which Permit Additional Tests of Hypotheses. Graphic analyses can be combined with appropriate experimental designs and used to rule out theories within the class of additive or interactive functions. One extension of factorial designs will provide a researcher with greater diagnostic capability for distinguishing between functions. A second extension of factorial designs provides a test of the form of the **J** function (Equation (3)) relating subjective to objective responses.

Designs that Vary the Amount of Information. In order to test between the unique predictions of the functions in TABLE II, experimental designs must vary the amount of information contained in the questionnaire items. A complete experimental design which varies the amount of information for three factors: Alert Forces, Airborne Forces, and Weather is shown in Figure 18. Every possible one-way and two-way factorial design is included along with the three-way factorial design. Each cell in the first three one-way matrices produces an item with just one piece of information. For example, an item containing one piece of information from the first one-way design might read: The status of 30 percent of your Alert Forces is known. An item

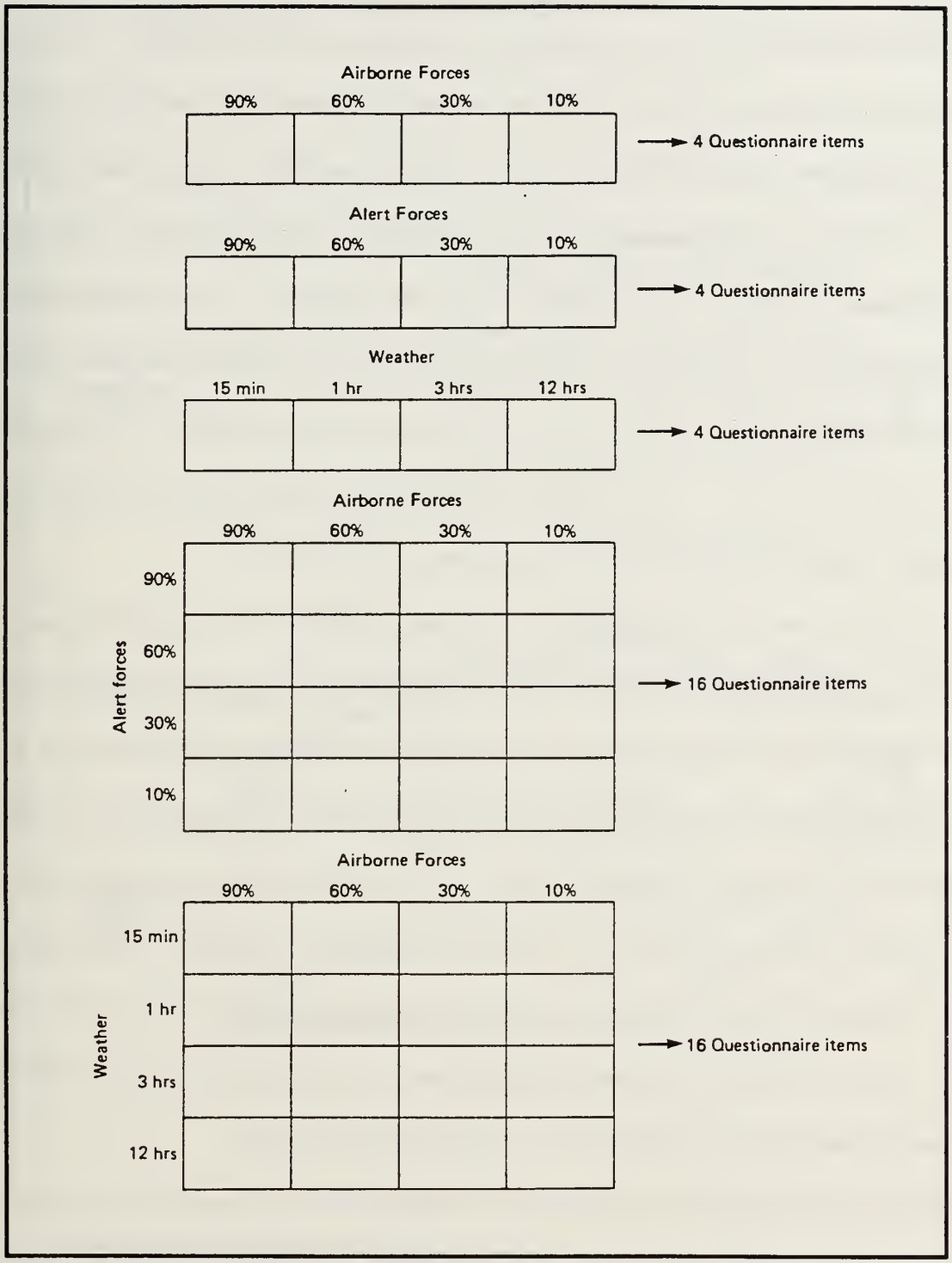

Figure 18. Experimental Design (for Three Factors) Which Varies the Amount of Information.[Ref. 11:p.22]

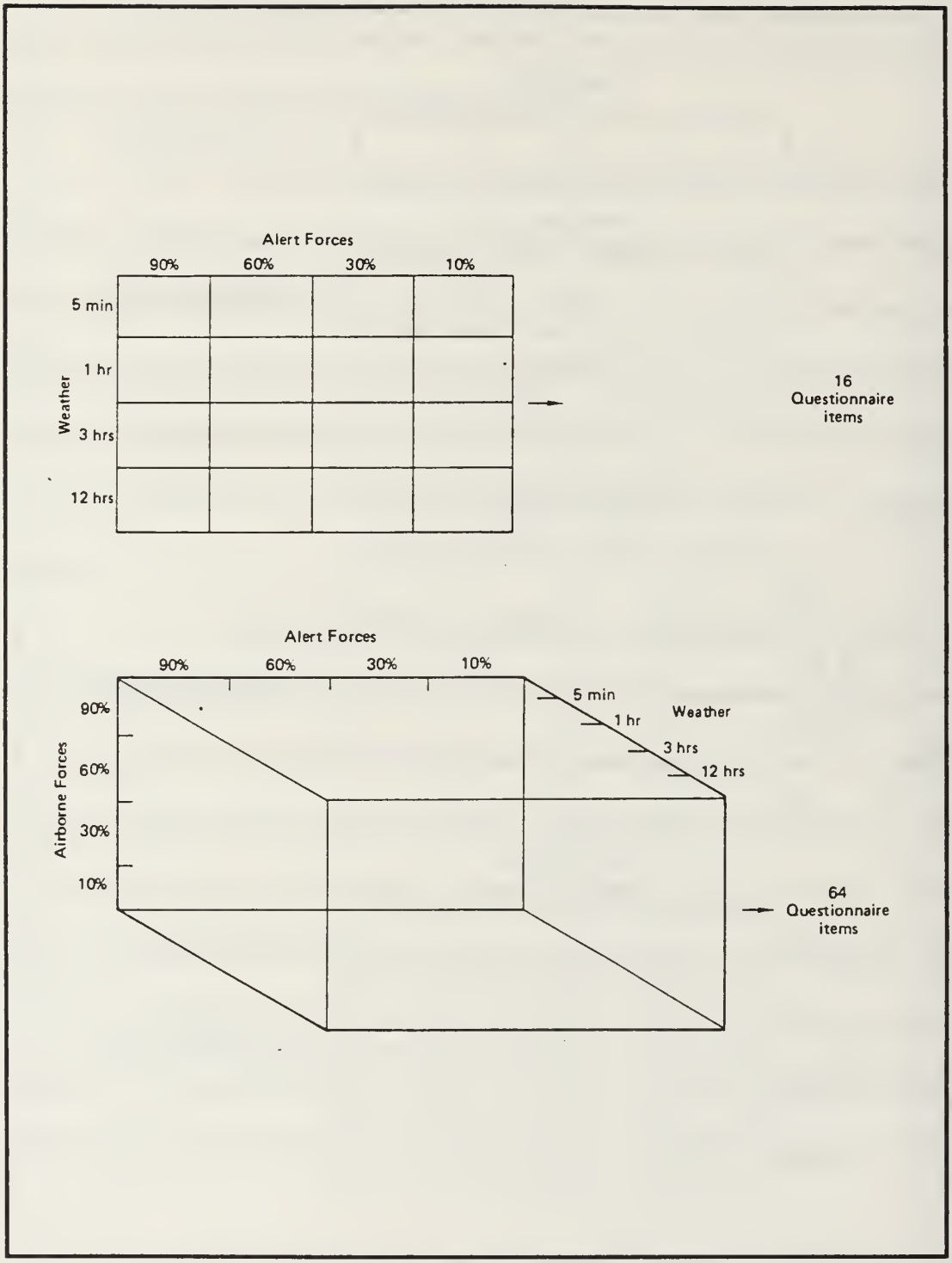

Figure 18. (Continued) Experimental Design (for Three Factors) Which Varies the Amount of Information.[Ref. 11:p.23]

from the third two-way matrix might read: The status of 60 percent of your Alert Forces is known. Reliable weather information is 15 minutes old. An item from the three way design might read: The status of 10 percent of your Alert Forces is known. The status of 60 percent of your Airborne Forces is known. Reliable weather information is one hour old. In this example a complete design for the three factors will produce 124 questionnaire items. Sometimes the expert respondents are instructed to assume a baseline level of capability for factors not presented, when items vary in the amount of information. Including this instruction depends on how reasonable the tasks are to the respondents if the instruction is omitted.[Ref. 11:p.21]

Graphing the hypothetical data shows how the design feature of varying the amount of information assists in diagnosing between the functions. By obtaining data for the entire design (Figure 18), a test between a multiplicative function and a range function would be possible if the data were interactive as in Figure 17A. Also, it would be possible to distinguish between the additive and averaging functions if the data were noninteractive (Figure 17B). Figure 19 shows the different predictions of these models. The plots of the hypothetical data are the same as those in Figures 17A and 17B, except for the dashed curve. The data for the dashed curve could have been gathered from a one-way design of the factor plotted on the x-axis.[Ref. 11:p.21]

For each panel in Figure 19, the prediction of the function written in the upper left hand corner represented by the relationship between the dashed curve and the other curves in the figure. Data for the dashed curve would be gathered from a

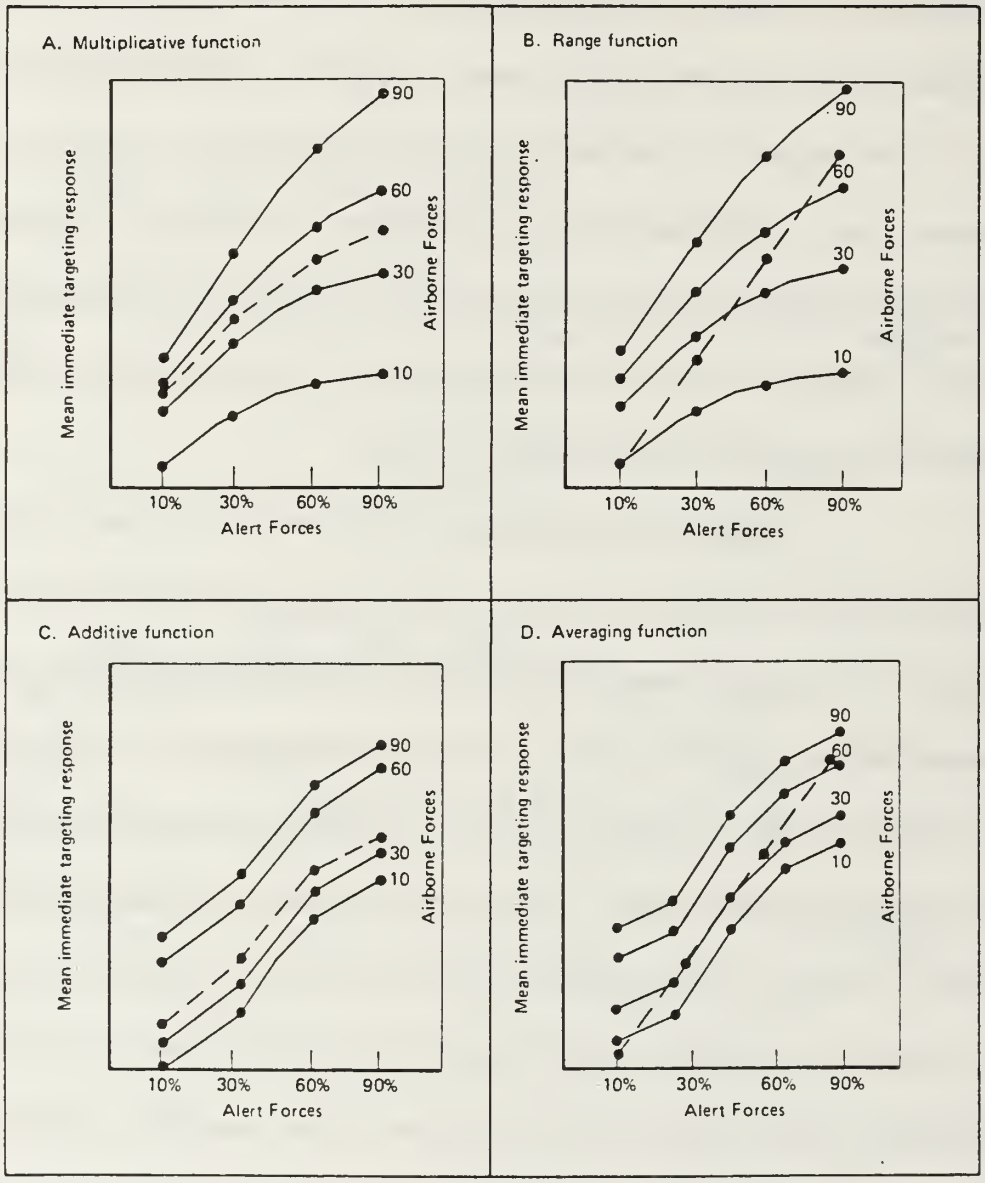

Figure 19. Predictions of Four Different Functions When Designs Vary the Amount of Information. (Dashed Line Represents Data for the Alert Force Factor Presented Alone.)[Ref. 11:p.24]

one-way design and data for the other curves, from a two-way design. The multiplicative function (Figure 19A) predicts that the dashed curve should follow the same increases or decreases in slope which would be expected from the family of curves. This prediction can be shown from the algebraic formulation of the multiplicative function. If two factors are provided for judgement, this model predicts the following form which the response should follow:

$$r = A_i B_j \qquad (4)$$

If a factor, A is presented by itself, the function predicts:

$$r = A_i x \qquad (5)$$

where x is the value of the missing information. The two values are multiplied and the resulting curve from only one piece of information should have a slope, which belongs to the slopes of the family of AB curves. An indication of the value of the missing information (i.e. the value associated with the missing information on Alert Forces) is provided by the height of the dashed curve (A alone).[Ref. 11:p.23]

If we assume the appropriate function to be the range function (Figure 19B) than the dashed curve should have a slope which is steeper than any of the other curves. The algebraic form of the range function illustrates this. If two factors, A and B, are provided for judgement, then the function predicts the following form for the response:

$$r = \frac{W_0 S_0 + W_A S_{A_i} + W_B S_{B_j}}{W_0 + W_A + W_B} + \omega\,(S_{max} - S_{min}) \qquad (6)$$

If only one factor (i.e. A) is provided for judgement, then the weight of the missing information goes to zero and the response should be in the following form:

$$r = \frac{W_0 S_0 + W_A S_{A_i}}{W_0 + W_A} \qquad (7)$$

Since, the denominator has decreased, the slopes of A should increase. This is shown by the dashed curve in Figure 19B. An in the previous case, the value of the missing information is indicated by the height of the dashed curve. For distinguishing between an additive and an averaging function, the same reasoning can be applied. Figures 19C and 19D show the different predictions of the additive and averaging functions.[Ref. 11:p.25]

Stringent tests among these four functions are provided in the design shown in Figure 18. The reason the tests are stringent is attributed to the repeated opportunities offered by the design for the predictions of a proposed function to fail. To determine if an interactive interpretation is appropriate for factors A and B, an interaction between them should be observed in the two-way and the three-way designs. Suppose the appropriate interactive function is the multiplicative function. A graph with A on the x-axis with a separate curve for each level of B together with A alone should resemble the graph with B on the x-axis with a separate curve for each level of A together with B

alone. In this case they should look like Figure 19A. Now, if interactions are observed among three factors (A,B,C) and a multiplicative function is appropriate, then the form shown in Figure 19A will result for all pairs of factors, including graphs of three factors (i.e. the AB design plotted on the x-axis with a separate curve for each level of C together with AB alone). These plots should resemble the forms shown in Figures 19B, 19C, or 19D for a range, additive, or averaging function, respectively.[Ref. 11:p.25]

When the design in Figure 18 is used, other graphic diagnoses of the data are possible. An example of this is the additive function (TABLE IIA), which predicts the effect of a factor should be independent of the factor(s) with which it is paired. This prediction can be evaluated graphically for each separately. This is shown for a factor, A, in Figure 20A. Data which would be obtained from the AC design averaged over C is represented by the top curve. Data which would be obtained from the ABC design averaged over B and C is represented by the next lower curve. Data which would be obtained from the AB design over B is represented by the bottom curve. The additive model (TABLE IA) j predicts all the curves will have the same slope.[Ref. 11:p.25]

An averaging model, such as the relative-weight model (TABLE IIA), predicts the effects of a factor depend on the number of other factors it is paired with. Therefore, the slopes of the same curves should vary by the amount of information contained in the item. This is shown by the curves in Figure 20B. A different factor is represented by each set of four curves. The curves in panel 1 of Figure 20B represent

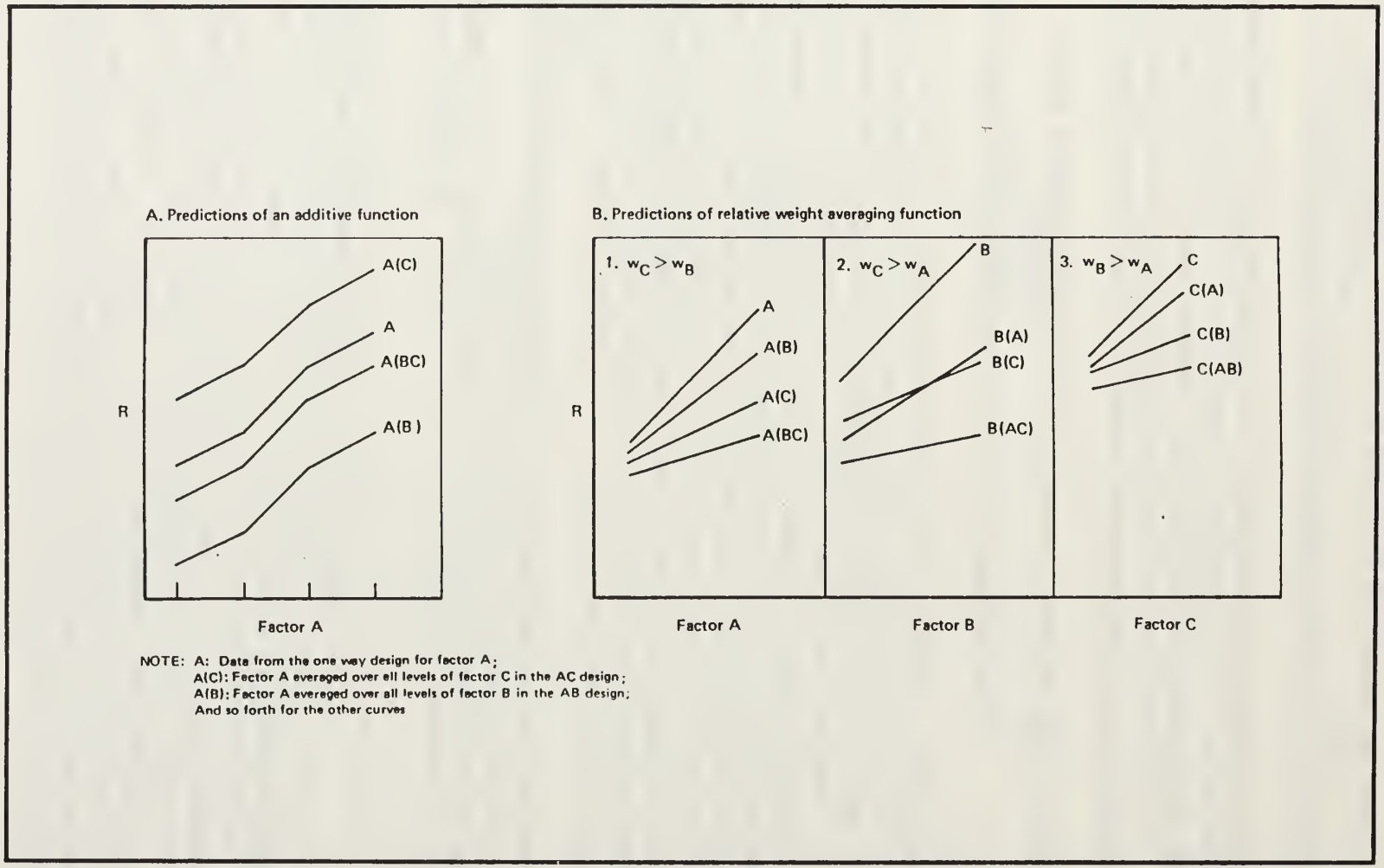

Figure 20. Graphic Preditions of Additive and Averaging Functions. (Examples are for Three Factors A, B, and C shown in Figure 18.)[Ref. 11:p.26]

the data from the four design in which factor A is included (i.e. A alone-the top curve, AB, AC, and ABC designs). In the same way, the curves in panels 2 and 3 of Figure 20B represent data from B alone, AB, AC, ABC, and C alone, CA, CB, CAB designs, respectively. The interactions and the order of the weights of the averaging function are revealed by the slopes of each set of curves. In panel 1 of Figure 20B, the curves suggest the weight of factor C is greater than the weight of factor B, since its slope is less. By looking at the functional form of the relative-weight function this prediction can be observed. Equation (8) is the slope of A when paired with only C.

$$\frac{w_A}{w_0+w_A+w_C} \qquad \textbf{(8)}$$

This results from the judgement which would follow the following model (Equation (9)).

$$r=\frac{w_A s_{A_i}}{w_0+w_A+w_B+w_C}+\frac{w_0 s_0+w_C s_{C_k}}{w_0+w_A+w_C} \qquad \textbf{(9)}$$

Note the difference when the slope of A is paired with B (Equation (10)).

$$\frac{w_A}{w_0+w_A+w_B} \qquad \textbf{(10)}$$

The difference between the two slopes in Equations (8) and (10) is the weight of factor C (Equation (8)) versus the weight of factor B (Equation (10)). Applying the same reasoning, the weight of factor C can be shown to be greater than the weight of factor A

from the order of the curves in panel 2 of Figure 20B, and the weight of factor B is greater than the weight of factor A from the order of the curves in panel B3. Therefore, a test of transitivity of the values of the weighting parameters in the averaging models is permitted by the experimental design shown in Figure 18.[Ref. 11:p.27]

The purpose of experimental designs is to test the unique predictions of algebraic functions, which are being analyzed. A function increases its credibility as the appropriate function to explain the data as the predictions of a particular function are observed repeatedly in the data. A function loses its credibility and would be eventually rejected as an appropriate theory to model the data, if the data do not follow the predictions of a hypothesized function.[Ref. 11:p.27]

Scale-Free Designs: Tests of the J Function. The scale-free design was developed to resolve the measurement problem of determining when it is appropriate to transform observed interactive curves, like those in Figure 17A, to parallel curves, like those in Figure 17B. The scale-free design is a method which provides a basis for deciding if observed interactions reflect the underlying subjective process (T in Figure 15).[Ref. 11:pp.27-28]

The importance of this design feature can be seen by referring back to TABLE II. If the interactions observed in the data are interpreted as perceptual, then the tests would be between interactive functions (TABLE IIB). Also, subjective values would be obtained from the interactive function, which best explains the data. But, if the interaction was transformed away, then the tests would be between noninteractive functions (TABLE IIA). In addition, a different function and set of subjective scale

values would be needed to explain the transformed data. Obviously, both interpretations of the data cannot be correct. The scale-free test provides a testable basis for determining whether or not to transform away observed interactions.[Ref. 11:p.28]

The basic concept behind the scale-free test is that by embedding two combination processes into one task, the interaction between the factors for one of the combination processes can be tested without making any assumptions on the form of the functions relating observed to subjective responses (J in Equation (3)) for the process. An example of this showing how the factors of emitter location and emitter coverage (Figure 12) combine additively in the combination process is explained in detail in *Introduction to the Subjective Transfer Function Approach to Analyzing Systems* [Ref. 11].[Ref. 11:p.29]

In summary, the major advantage of using a comparison task in the scale-free design is that a subtractive function has performed well in accounting for comparison judgements for a variety of judgement dimensions. Another task or function could be included. Data could be transformed in accordance with the predictions and the scale-free values obtained from the function, as long as the data were monotonically related to psychological response to perform a scale-free test of the embedded combination process. Because the scale-free design requires embedding factorial designs, the length of questionnaire increases substantially when more factors are added. The next section discusses the guidelines for decreasing the number of items, while maintaining sufficient constraints for testing proposed STFs.[Ref. 11:p.30]

b. Selecting Experimental Designs

It is important to obtain information on interactions among factors by using the scale-free design by varying the amount of information contained in each item. This is done in order to test between the alternative functions, when system factors are being determined for the first time and a decision must be made to determine the appropriate STF and subjective values. When a lot of factors are being analyzed, such as the six factors in experimental unit 1 (Figure 12) and each has three to four factor levels, it becomes practically impossible to field a questionnaire, which includes a set of items from these designs.[Ref. 11:p.31]

A combination of two techniques makes it possible to come up with a reasonably sized (i.e. up to 200 items) questionnaires. The first technique is gathering judgement data in stages. This requires *pilot studies* to be conducted before the final STF testing stage and provides information on the effects of the factors on judgements. The second technique is selecting a subset of a complete array of experimental designs. This technique focuses on an experimental design selection process to reduce the length of the questionnaires, while keeping the necessary constraints to adequately test among the proposed STFs.

(1) Pilot Study Phase. A preliminary investigation of the system factors being considered aids greatly in the following areas: 1) assessing if the judgement tasks are feasible to the respondents, 2) reducing the number of structure/function hypotheses to be tested in the STF testing phase, and 3) providing a stronger foundation for conclusions concerning appropriate STFs through repeatability of

the results. During the pilot study phase, a number of areas can be stressed. Emphasis could be placed on the main effects of factors, distances between factor levels, scale-free interactions between the factors, and testing predictions of the STFs being investigated. Since, the amount of information obtained in this phase is dependent on time, resources, and the availability of the respondents, different questionnaires, each addressing a different question about the factors, may be fielded to two or three respondents within an expert group. The following reasons may require more than one questionnaire for an experimental unit be fielded to ensure adequate determination of experimental questions: 1) questions are generated by the results from a first fielding, which require answering, 2) results of a questionnaire are unclear, or 3) more information is required about the relationship among the factors.[Ref. 11:p.31]

Guidelines for a first round of questionnaires are provided in the following description of experimental designs. All factors and factor levels should be included in some aspect of the experimental design. This could be done by including all two-way factorial designs or a mixture of two-way and three-way designs. This will depend on the size of the questionnaire, which is generated. In order to get an idea of higher-order interactions, four-way and five-way designs might be included in the design. The number of factor levels could be reduced by two for these larger designs. It is a good idea when reducing the number of factor levels for a design to include in the selection what the highest and lowest values for each factors to ensure the full range of the factor dimension is covered. For some factor pairs scale-free designs could be

included. The number of respondents available will determine how much can be fielded in each round.[Ref. 11:p.31]

The results from the first round of questionnaires may suggest changes in the factors, which require more data collection. First, the preliminary results may point to a factor which should be redefined. Confusion about the definition of a factor, no effect of the factor on judgements, or individual differences in the effects of the factor on judgements are all indications which may occur. Second, the preliminary results may reveal little difference between some of the factor levels. New levels may have been hypothesized. If so, they should be tested. Third, the factors displaying interactions should be refielded in scale-free designs, if this wasn't done during the first round. Some of these factors could be embedded in three-way designs depending on the size of the questionnaire generated by the design and the number of available respondents.[Ref. 11:p.32]

The shape of the experimental designs used in the STF testing phase are guided by the results from the pilot study. If a factor had no effect on judgements it is no longer included in the experimental design. Factor levels close in value are replaced by one level. Scale-free tests provide information about whether factors will combine interactively or noninteractively and thereby reduce the number of STFs needed in the final design.[Ref. 11:p.32]

(2) The STF Testing Phase. During this phase, it is important to use experimental designs, which allow for adequate tests of the STFs being considered in each experimental unit. This approach will lead to the final conclusions about which STFs are

appropriate. As shown in Figure 18, a major experimental design feature varies the amount of information in an item. Assume between three and five levels for each factor. If more than three factors are included in an experimental unit, a complete design of all the possible number of factors would produce a questionnaire too long to field. Therefore, it is necessary to choose a subset of factorial designs from a complete array of designs, which allows for sufficiently stringent tests of the STFs being considered. The selection of factorial design subsets should be primarily based on knowledge of the unique predictions of the most viable STFs. These STFs would be based on the information from the pilot study. Guidelines for selecting or excluding design subsets is provided below.[Ref. 11:p.32]

1. For larger designs (i.e. three-way and larger) reduce the number of factor levels for those factor combinations, which received more stringent test (i.e. all factor levels were used in the pilot study). Retain the top and bottom levels of the factor.[Ref. 11:p.32]

2. In experimental units comprised of four or more factors, use the strategy of confounding factors. This means not crossing each factor fully with every other factor. However, generate items which include a factor level of each factor. Ensure factors confounded in larger designs are unconfounded in smaller designs. Use the pilot study data to determine which factors to confound. Factors correlated in the real world might be chosen for confounding in the experimental design.[Ref. 11:p.32]

Instead of confounding factors, each can be reduced to two levels, the top and bottom levels and combined in a 2^n design, where n is the number of factors

in the experimental unit. If this allows for a reasonable questionnaire length, this approach provides much more information than confounding factors about the relationship among the factors.[Ref. 11:pp.32-33]

3. Provide adequate tests of proposed STFs by including enough subsets of one-way, two-way, etc. factorial designs. The adequacy of the tests are guided by the knowledge about the unique predictions of the STFs being considered. The researcher should include a subset of factorial designs, which provide a mathematically unique solution of the parameters of the STFs being considered. Thus, the subset of designs must produce a set of linear equations and unknowns for each STF from which a unique solution for the unknowns can be obtained. Additions should be guided by the knowledge about the unique predictions of the STFs being considered and the questionnaire length.[Ref. 11:p.33]

c. Collecting the Judgements Data

The collection of judgement data is comprised of selecting respondents and administering the questionnaire. The respondents associated with each experimental unit should have credibility to those requesting the evaluation of the system. In addition, their availability and professional characteristics need to be taken into consideration as these areas will impact the formulation of the structural hypotheses.[Ref. 11:p.33]

The administration of the questionnaires involves ensuring respondents are familiar with the task and the procedure for filling out the questionnaire. It is important to ensure respondents feel at home with the situations they will have to consider and understand the judgement task required of them. To ensure task

familiarization, the respondents are provided a preparatory session in which they are briefed on and allowed an opportunity to discuss any necessary background information, the factor definitions, the factor levels, and the judgement task. The respondent group's participation in structure development or pilot studies will determine the length of this session. After completion of the preparatory session, the respondents are provided 10 to 20 representative items to familiarize them with the task. Finally, they are provided the questionnaire to fill out.[Ref. 11:p.33]

3. Determine the STFs and the Final Structure

Data analysis is performed after each data collection session. These analyses provide tests among the competing STF/structural hypotheses. The structural hypotheses focus on the appropriateness of the hypothesized intermediary factors.

a. Data Analysis

(1) Determine Individual Differences. The first step in analyzing the data is looking at each respondent's data separately to observe the similarities and differences in the effects of the factors on the judgements. When there are no differences in the ordering of the factor levels and the interaction pattern of respondents' data within an experimental unit, the data is combined. If either the factor level ordering or the pattern of the data is different, then the two sets of respondents' data are different and neither set of data is combined for the analyses. However, one of the goals of the pilot study is to find factor definitions and task descriptions which affect the respondents in similar ways. This is because it simplifies the STF approach to have one STF at each

path in the structure. If the differences cannot be resolved, then research should be conducted to determine the source (i.e. military rank, training, etc.) of the differences and incorporate the differences into the STF theory.[Ref. 11:p.34]

(2) Testing Between the STFs. A proposed STF's explanatory power is its ability to reproduce the systematic details of the data. Two major steps for testing between the abilities of proposed STFs to accomplish this are performed in each experimental unit. The first step primarily uses graphic analysis to reduce the number of structure/function hypotheses down to a select few. The second step uses additional graphic tests in conjunction with the least-squares data-function discrepancy criterion to test among the remaining hypotheses. These two steps are discussed below in more detail. For this discussion, assume scale-free tests are fielded in the pilot study stage and have provided the basis for assuming the function, J (Equation (3)), is linear.[Ref. 11:p.34]

Reduction of the Number of Possible Hypotheses by Graphic Tests. Graphic tests provide a powerful diagnostic tool for distinguishing between algebraic functions as shown in Figures 19 and 20. Using this tool the number of viable hypotheses can be drastically reduced. Usually in large designs (four factors or more) a few hypotheses are retained as possible explanations of the data, because many combinations of factors which would support a particular hypothesis may not have been fielded (in an attempt to reduce the length of the questionnaire). Thus, the more extensive the pilot study the fewer the number of ambiguities. The following examples are presented to illustrate this step.[Ref. 11:p.34]

In the first example, assume the experimental design has six factors (i.e. experimental unit 1 in Figure 12), labeled A, B, C, D, E, and F. If the pattern of the data for factors A, B, C, and D for the design fielded is similar to that illustrated in Figure 21B, but the convergence is smaller and the data pattern for factors E and F are similar to the pattern in Figure 21C, then a viable hypothesis is a positive ω weight of the range term for factors A, B, C, and D and a multiplicative function for factors E and F. If a small convergent interaction for factors A, B, C, and D leads to the retention of the relative-weight averaging function for these factors, then the question of how to combine these two groups must be answered. If a graph of EF (x-axis) with a separate curve for each ABCD combination is similar to Figure 21C, then a multiplicative function is a viable hypothesis. The combination functions for the two factor sets is illustrated in Figure 22. Also, depicted is the multiplicative combination function for combining the two sets. Note, two STFs are retained for the ABCD factor set.[Ref. 11:pp.35-36]

For the second example, the experimental unit has five factors, A, B, C, D, and E. Assume graphic diagnostics revealed a divergent interaction (Figure 21A) for factors C, D, and E and that the interaction is small. In this case the data suggests the retention of both a relative-weight averaging function and a range function with a negative ω. Now assume the graphic diagnoses also reveal a multiplicative (Figure 21C) relationship between factors A and B and additivity (Figure 21D) for the two factor sets. A graph of the AB (x-axis) with a separate curve for each CDE combination is similar to Figure 21D. One way of grouping the factors is illustrated in Figure 22B, with two possible functions for the CDE factor set. Since, all interactions in the data were

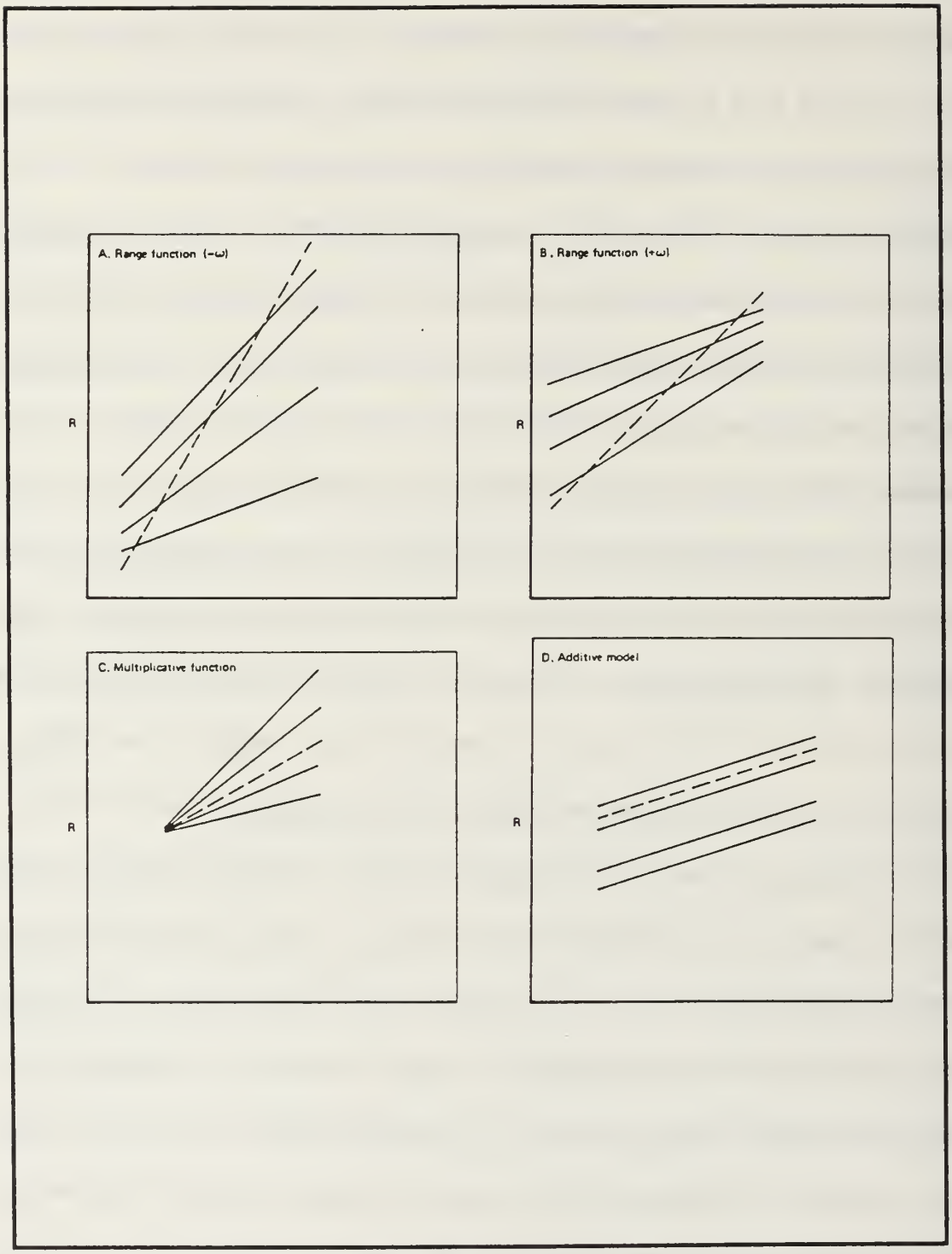

Figure 21. Hypothetical Data Predicted by Different Algebraic Functions. (The Dashed Curves Represent the Data for the Factors Plotted on the X-Axis When Presented Alone.)[Ref. 11:p.35]

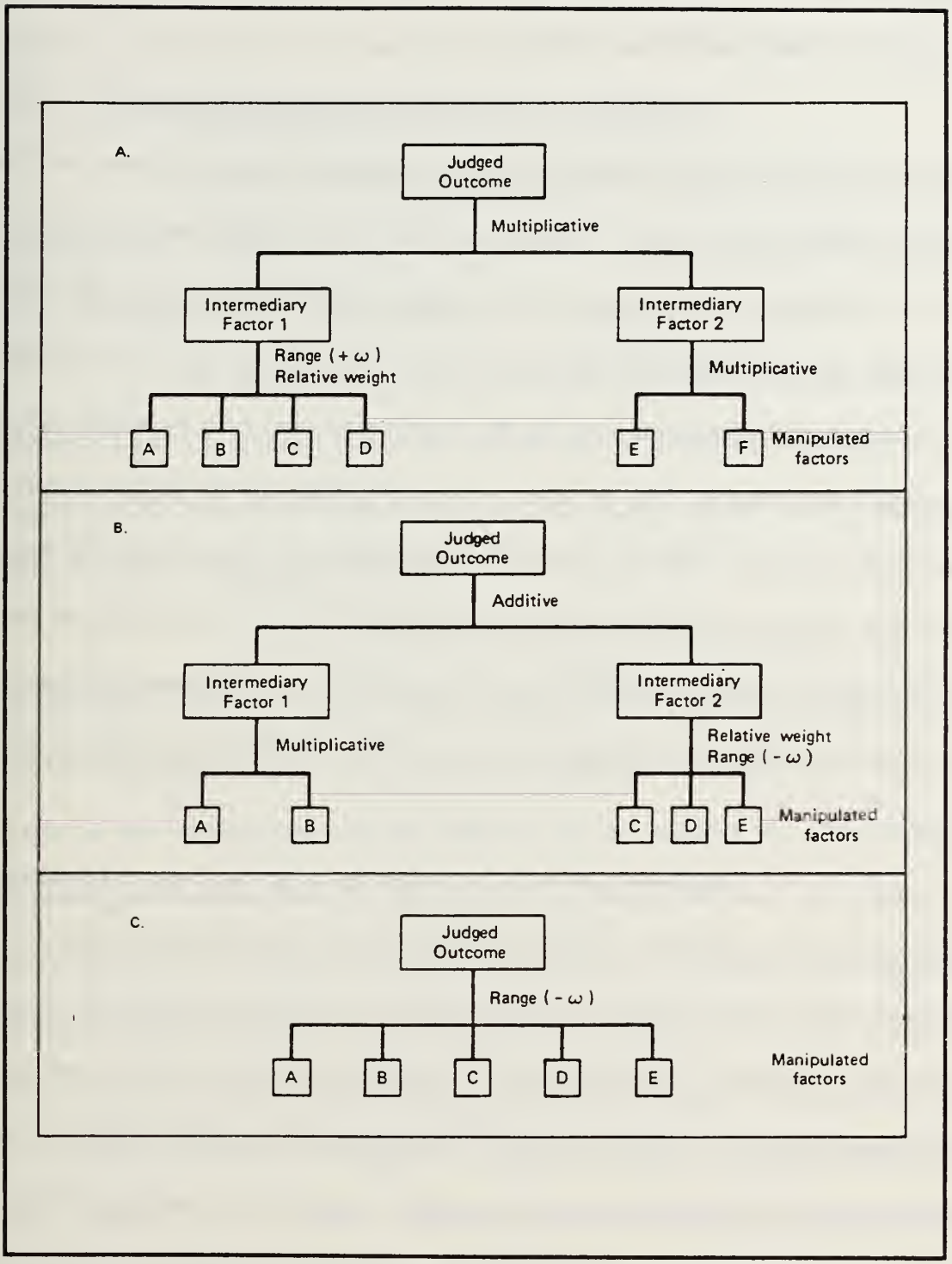

Figure 22. Alternative Structure/Function Combinations.[Ref. 11:p.37]

divergent, an alternative approach would be to investigate a five factor range model with a negative ω (Figure 22C).[Ref. 11:p.36]

Testing Between Selected Structure/Function Hypotheses. Tests between structure/function hypotheses for each experimental unit are done with the STEPIT program [Ref. 11:p.36]. This parameter-estimation program selects parameters, which minimize the sum of squares discrepancies between the data and the STF's predictions. By embedding STFs associated with intermediary factors in to the function at the outcome or suboutcome path for the experimental unit, each hypothesis can be written into the program. For example, a structure/function test for the five factors in Figure 22B could be to embed two functions: a multiplicative function for the AB factors and a range function. Another example is a structure/function test for a five factor range function (Figure 22C). The STEPIT program provides the sum of squares data/function. The "best-fit" STF for an experimental unit is the STF with smallest discrepancy.[Ref. 11:p.36]

As previously discussed, when deviations are large and systematic for the statistically best function, the function would be rejected as the appropriate STF. Consider graphs which plot both the predicted values (r in Equation (2)) and the obtained values (R in Equation (3)) on the y-axis for the different factorial designs used in the experimental designs. They provide the ability to assess the magnitude, direction, and systematic nature of data/function deviations. These graphs help in deciding whether or not to reject the functions and in determining a "correct" function. If the pattern of

deviations produce a function which cannot be adequately tested on the available data, then the experiment needs to be redesigned and new data collected.[Ref. 11:p.36]

b. The Final System Structure: An Example

The structure of the system within each experimental unit is determined by the fit of the STFs to the data. When an STF has been determined, the subjective values associated with the system factors will be known.[Ref. 11:p.38]

The final structure shown in Figure 23 resulted from the hypotheses (Figures 12 and 13 represent two of the hypotheses). The difference between the final structure and the hypothesized structure (Figure 13) is the omission of the intermediary factor, Execution Status.[Ref. 11:p.38]

At each path a function is chosen as the appropriate STF. At the top of the final structure, a range function with a negative ω term best accounted for the data of the immediate targeting respondents. Divergent interactions found among the six factors in this experimental unit are indicated by the negative ω. This finding is interpreted to mean the better the capability on one factor, the more of a difference it makes how good their capabilities are on the other factors. The data of the targeteers at the target identification was best accounted for with a range model with a positive ω. The convergent interaction indicates the greater the knowledge of the targeteers about enemy emitters, the less of a difference it made how much information they had on enemy vehicles and vice versa. Vehicles and Emitters, the two intermediary factors (Figure 13) were kept in the final structure. At the vehicle path, the range function with

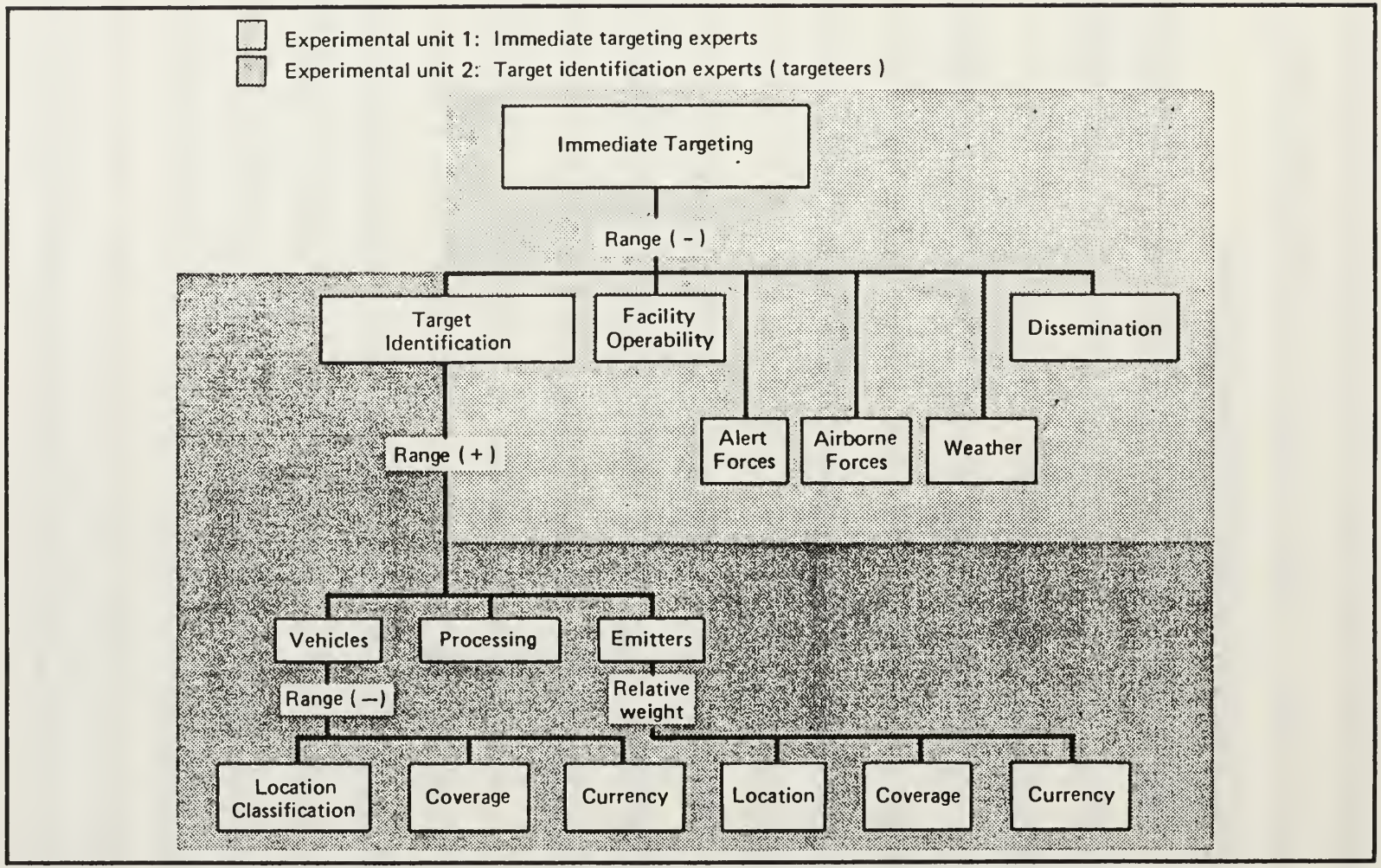

Figure 23. Final Structure and STFs for Immediate Targeting Structure (Figure 12).[Ref. 11:p.40]

a negative ω term best accounted for the divergent interactions displayed by the three factors in this experimental unit. At the emitter path, the relative weight function with an initial impression (TABLE IIA) best accounted for the emitter data. The overriding trend in these data was independence between the factors on the target identification judgements.[Ref. 11:p.38]

4. Evaluate Systems

The final step in the STF approach is the evaluation of the system. Since, the system's final structure and STFs have been determined it is possible to evaluate systems which differ in their capability levels. The capabilities are defined in terms of the system's primitive factor levels. Systems are different when they differ in at least one of the factor levels. In Figure 22, location/classification, coverage, and currency are the factors of the experimental unit, Vehicles; location, coverage, and currency are the factors of the experimental unit, Enemy Emitters. The primitive factors in this structure are those factors identified above and Processing, Facility Operability, Alert Forces, Airborne Forces, Weather, and Dissemination.[Ref. 11:p.41]

The outcomes and suboutcomes of a system are determined by computing the STFs at the primitive factors paths, transferring the computed values for each output to the STFs with which they are linked and continuing this process until all the STFs have been computed and all suboutcomes and the final outcome(s) have been computed.[Ref. 11:p.41]

To begin this process, the subjective input values to the STFs at the primitive factor paths are needed. When the STEPIT program tests the STF, it prints out all of the

parameters values (i.e. factor weights, stimulus scale values, and the value of the range function term, ω). Note, a scale value is not available for a primitive factor level (stimulus) that was not used in the experimental designs. If the primitive factor is defined along a physical continuum (i.e. time, distance, percent), then the functional form of the function, H (Equation (2)) can be obtained by plotting subjective values associated with the manipulated factor levels as a function of the physical values. The factor levels within the range of manipulated factor levels can be obtained from the curve which best fits these points. If the primitive factors are written descriptions, then the resulting plot is a set of points which cannot be connected. If a system is defined by a different description then the experiment needs to be refielded to obtain a scale value for the new primitive factor level.[Ref. 11:p.41]

An example of this evaluation process is shown in Figure 24. Assume the systems shown in Figures 24A-C are being compared on how well the immediate targeting people thought they could do their job or the percent of force applications they thought they could exploit. Circled primitive factor levels define the different systems.[Ref. 11:p.41]

First, the subjective values associated with these factor levels are obtained from the questionnaires completed by the expert respondents. The psychophysical functions for the vehicle coverage and currency factors and subjective points for the location/classification factor are shown in Figure 25. The projected subjective values are those needed for the range function (see top of Figure 25) to compute the vehicle intermediary factor. In the same way, the scale values for primitive factor levels at other

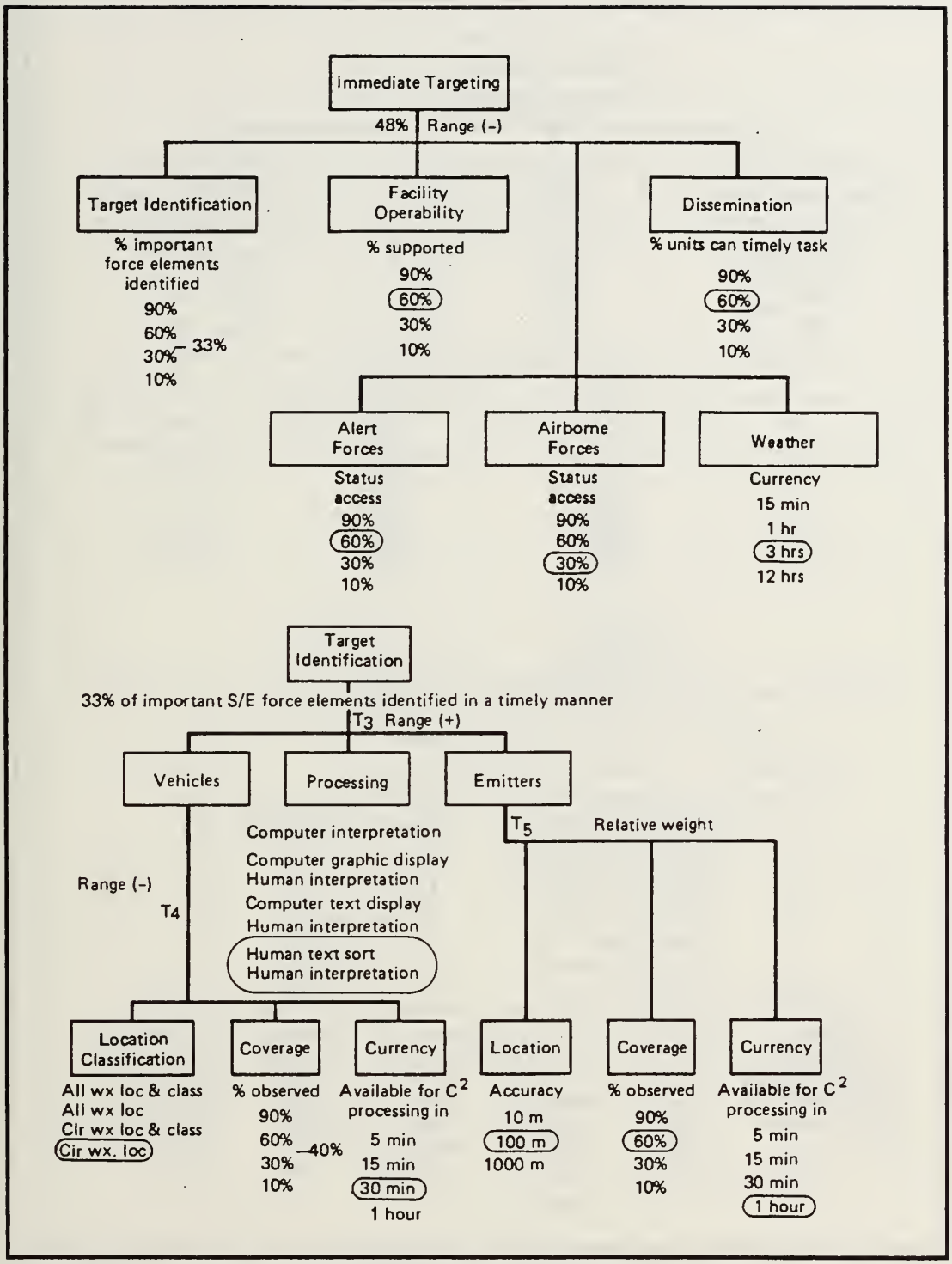

Figure 24A. System Comparisons.[Ref. 11:p.42]

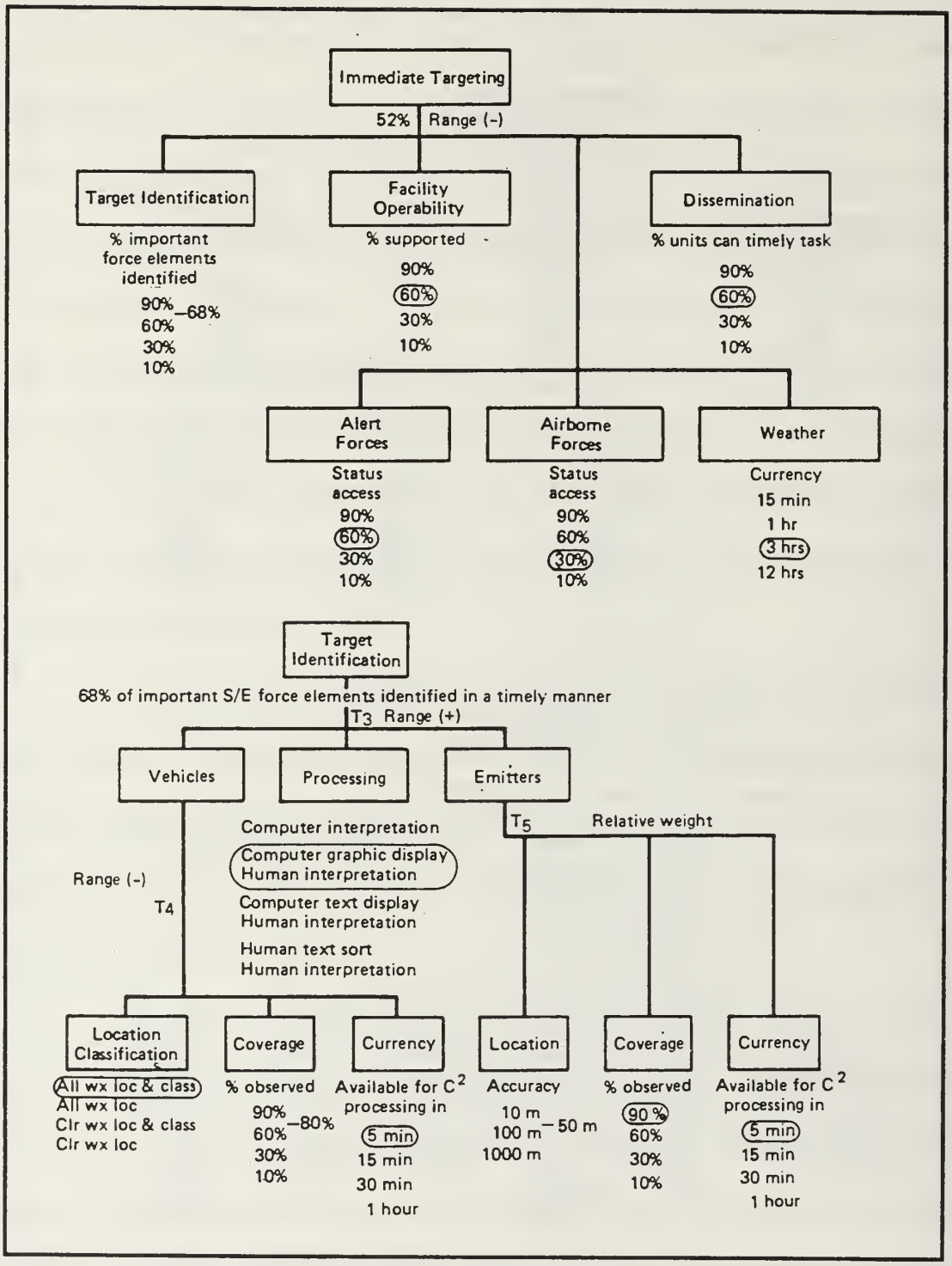

Figure 24B. System Comparisons.[Ref. 11:p.43]

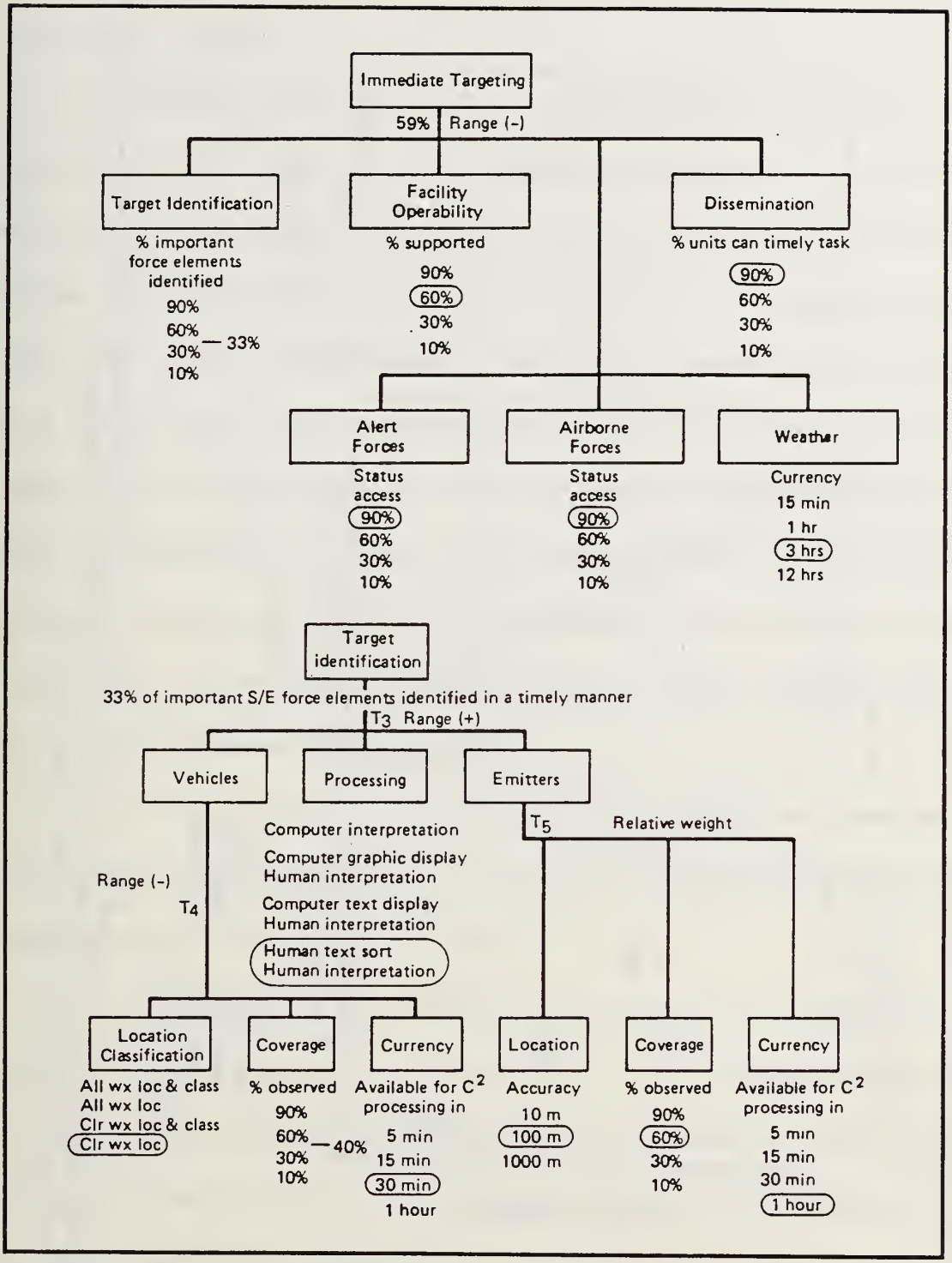

Figure 24C. System Comparisons.[Ref. 11:p.44]

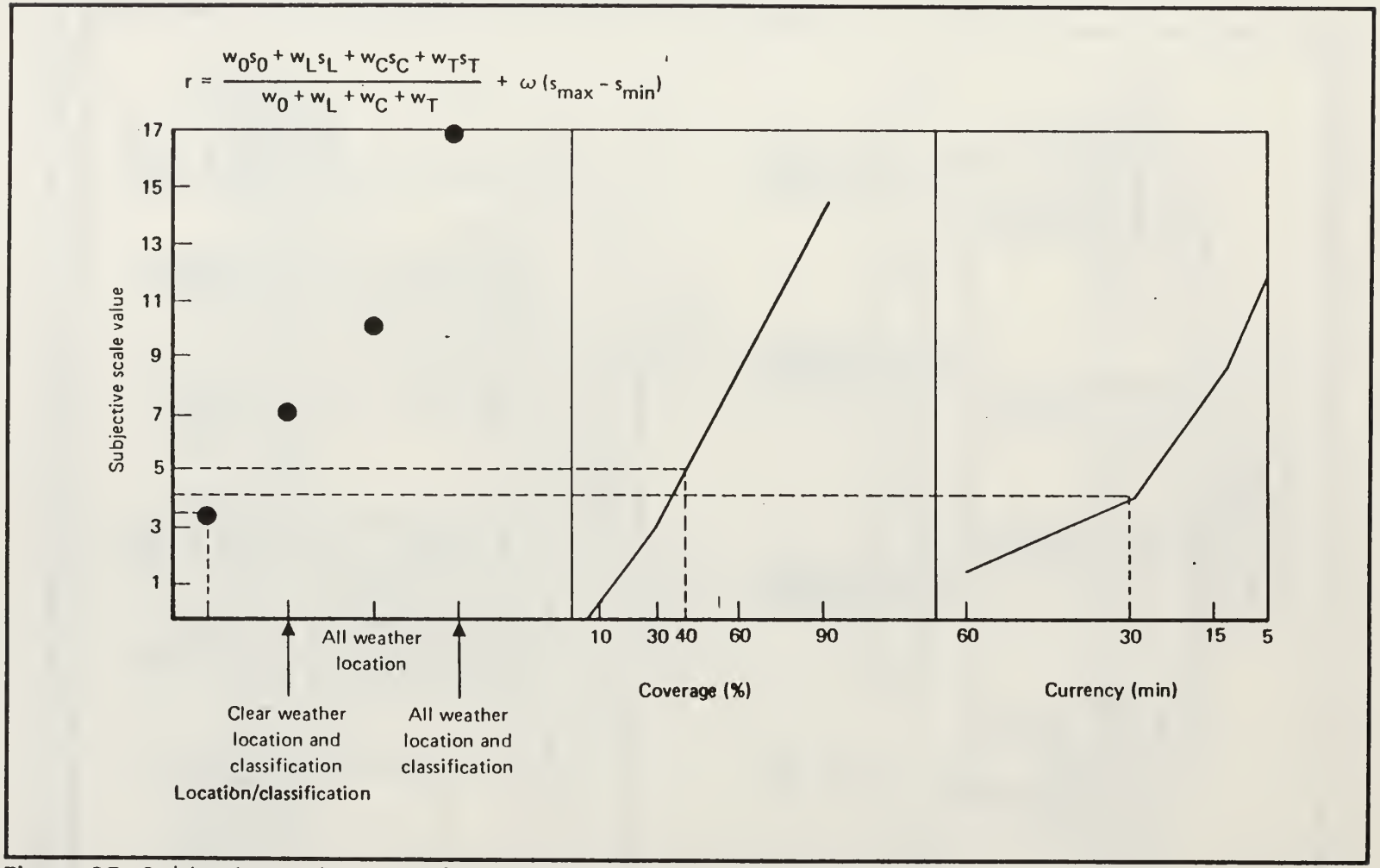

Figure 25. Subjective Scale Values for the Location/Classification, Coverage, and Currency Factors Shown in Figure 12.[Ref. 11:p.45]

paths can be obtained. This computation procedure is continued to the top of the structure.[Ref. 11:p.41]

The STFs in the system shown in Figure 24A predict the targeteers would perceive they could identify approximately 33 percent of the important targets. This is used as the target identification factor level. From this immediate targeting respondents perceive they could exploit about 48 percent of the important immediate targeting opportunities. In Figure 24B, the ability of the targeteers to identify targets is increased to 68 percent and the other immediate targeting capabilities are kept at the levels in Figure 24, which means the ability to do immediate targeting increases to 52 percent. In Figure 24C, the system capabilities are set at the levels shown in Figure 24A and the immediate targeting capabilities are improved for Alert Forces, Airborne Forces, and Dissemination, this results in increasing the ability to exploit immediate targeting opportunities to 59 percent.[Ref. 11:pp.45-46]

The primitive factor levels selected for the evaluation are determined by considering the systems being evaluated for purchase, production, development, or present capability levels, to name a few.[Ref. 11:p.46]

The tradeoffs for the contribution of two factors to an outcome can be assessed by looking at a graph of STF predictions. This is depicted for the facility operability and dissemination factors (Figures 24A-C) in Figure 26. Notice that the subjective judgement on the y-axis is approximately the same for a dissemination level of 10 percent and a facility operability level of 90 percent as for a dissemination level of 60 percent and a facility operability level of 30 percent. Other tradeoffs between these

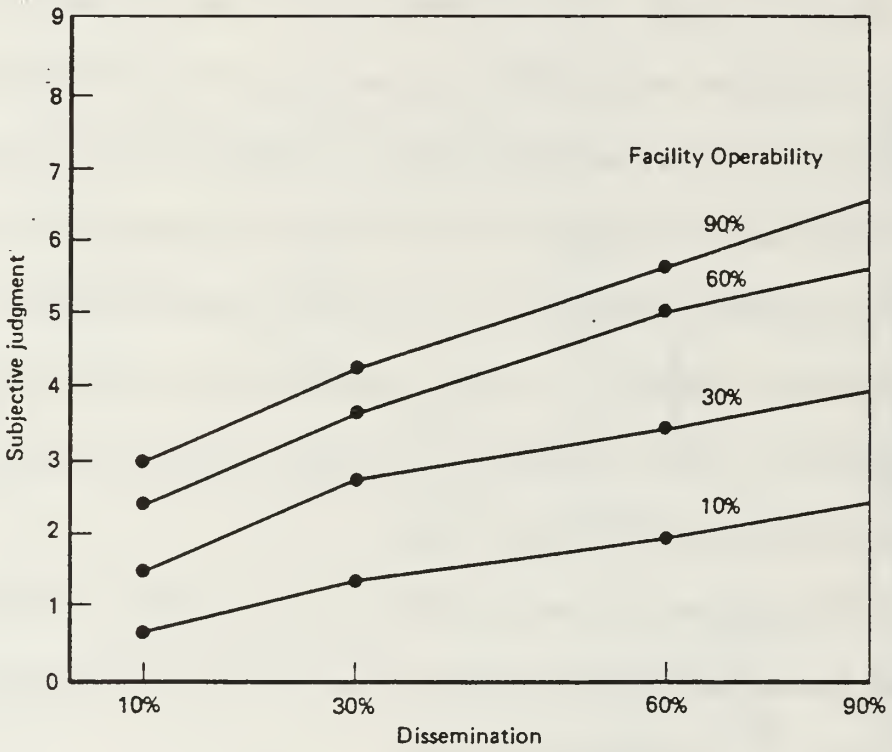

Figure 26. Theoretical Predictions.[Ref. 11:p.46]

factors can be assessed by drawing horizontal lines through the theoretical curves. To assess the tradeoffs between three factors, a graph for two factors (Figure 26) can be plotted at each level of the third factor. To assess the tradeoffs between four factors, a graph for every factor level combination of two of the factors on the x-axis can be plotted with a separate curve for the third factor, and a separate curve for the each level of the fourth factor. Evaluating tradeoffs with graphs is useful when the decision about which system changes to make involves only a few system factors.[Ref. 11:pp.46-47]

In summary, the STFs (when computed) provide measures for each suboutcome in the structure making it possible to see where changes are occurring. This provides valuable information to those people who have to determine what systems to develop or purchase. Specifically, it provides systematic information about the effect of changes and where the effect occurs in the system.[Ref. 11:p.47]

V. APPLICATION OF THE STF APPROACH TO THE EVALUATION OF TACTICAL AIR C^2

A. OVERVIEW

Application of the STF approach to evaluate the MOE of tactical air C^2 and its impact on the MOFE of tactical air forces is the focus of this chapter. Realistically, a full implementation of the STF approach is beyond the scope of this thesis and the time, manpower, and analytic capabilities available to the author. Application of the first step of the STF approach should be sufficient to show it's value as a viable method for relating the levels within the evaluation hierarchy (i.e. MOE and MOFE). The example used is taken from work done by The Rand Corporation in the evaluation of tactical air C^2 on the employment of tactical air forces. Working in conjunction with Air Force personnel from the Tactical Air Command, the Tactical Air Forces Interoperability Group (TAFIG), and Headquarters Air Force, Studies, and Analysis, Rand examined the contribution of enemy information to the effectiveness of C^2 in employing tactical forces.[Ref. 6:p.15]

Chapter V will look at how steps one, two, and four of the STF approach can be applied to analyzing the relationship between an MOE of a system and the MOFE of a force which it supports. Step one, will present a complex system representation for tactical air C^2 and tactical air forces. It will look at the structural hypotheses based on the information presented in Chapter III on tactical air forces and tactical air C^2. System

outcomes and system factors will be identified as well as an alternative structure. This will be followed by a discussion of STF hypotheses. Step two will look at possible designs for judgment experiments. Finally, step four will look at evaluating the representation for tactical air C^2 and tactical air forces.

B. DEFINING A COMPLEX SYSTEM REPRESENTATION FOR TACTICAL AIR C^2 AND TACTICAL AIR FORCES

The first step in the process of defining a complex system representation for tactical air C^2 and tactical air forces is to acquire information from a (pre-chosen) group of expert respondents identifying the important system outcome(s). The next step is to collect information identifying the components which could affect the system outcome(s). Now, a system's hierarchical structure can be hypothesized from the list of possible outcome(s) and components, which have just been acquired. Some of the components are hypothesized to be influenced by other components thereby providing an intermediary function in their impact on the final system outcome.[Ref. 7:p.13]

1. Structural Hypotheses

In Chapter III, tactical air forces and tactical air C^2 were decomposed into their functional elements. If this is considered as information collected from a group of expert respondents, then the following system outcome and system factors can be identified. A structural hypothesis for the hierarchical structure for tactical air C^2 can also be formed.

a. System Outcomes and System factors

For this complex system, employment of tactical air forces (air/land battle) is the overall (final) system outcome. The employment of tactical air forces is accomplished by flying or having the potential to fly military combat missions, which were identified in Chapter III as tactical air operations (TAOs). Three TAOs were determined as having the most impact on the employment of tactical air forces: close air support, interdiction, and tactical airlift. These are the system factors for the air/land battle and suboutcomes for tactical air C^2.[Ref. 7:p.27]

The effectiveness of tactical air C^2, will be viewed in terms of how it affects the military events in wartime, in this case, the TAOs. Thus, tactical air C^2 is assumed to affect military events through its impact on the performance of TAOs. As discussed in Chapter III, tactical air C^2 is composed of the following elements: planning, directing and controlling. Thus, the effectiveness of TAOs is based on how well they can be planned, directed, and controlled. The three TAOs are suboutcomes and the three elements of tactical air C^2 are the corresponding system factors for each suboutcome.

Planning, directing, and controlling are all made up of the following elements: friendly information, enemy information, processes (the way information is made available for use in the system) and communications (used to give directions to tactical air forces). Here planning, directing, and controlling are system suboutcomes and their elements are the system factors. Because these system factors are not decomposed any further, they are the primitive factors for this structural hypothesis.

Now that the system outcomes and factors have been identified as well as the causal relationship between them, a hierarchical structural hypothesis can be formed. This is shown in Figure 27. The relationship between the system outcomes and factors are stated as preliminary hypotheses in TABLE III. They are considered preliminary because these come before hypotheses, which specify the factor levels and combination models, which explain how the factor levels affect the judgement of the expert respondents. The development of a hierarchical representation is developed from a series of hypotheses about the causes and effects existing within a system affecting the overall system outcome.[Ref. 7:pp.13-15]

After the hypotheses are formed, the tactical air C^2 system can be divided into experimental units. Each experimental unit represents a causal relationship between one system (sub)outcome and its corresponding factors. In Figure 28, the tactical air C^2 system (Figure 27) is labeled with numbers representing the experimental units. These units correspond to the preliminary hypotheses listed in TABLE III. Each experimental unit contains the dependent (i.e. suboutcome) and independent (i.e. factors) variables necessary to test its hypothesis. Also, the numbers labeled on the structure in Figure 28 correspond to the STF hypotheses, which will be discussed in a later section.[Ref. 7:p.15]

b. Alternative Structures

Alternative structural hypotheses result when different preliminary hypotheses exist about how the expert respondents combine information included in a system's capabilities. As discussed in Chapter IV, they can occur during the development of the structure with the respondents or during the analysis of the data.

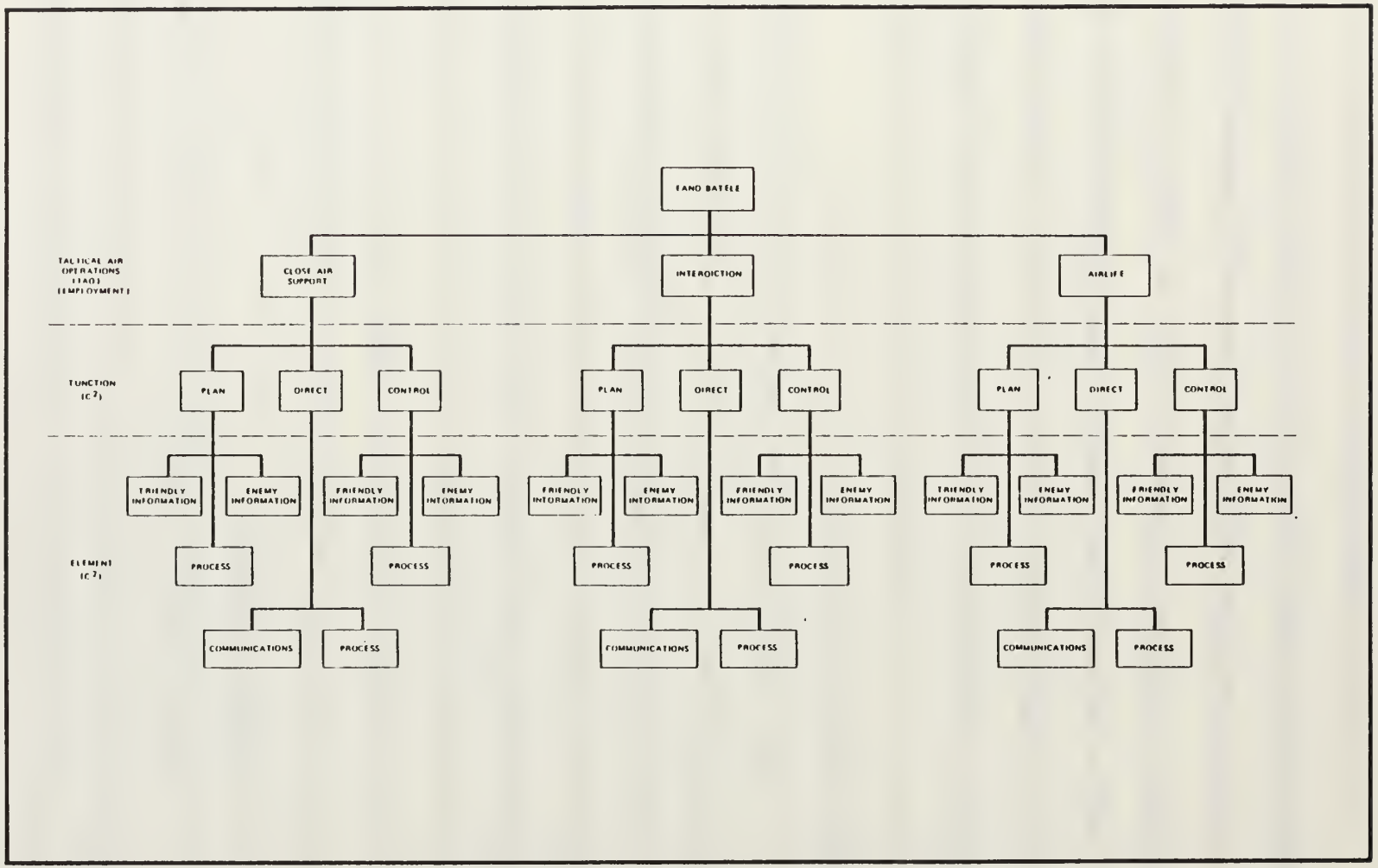

Figure 27. Tactical Air C2 and Tactical Air Force Representation.[Ref. 7:p.3]

TABLE III. PRELIMINARY HYPOTHESES ASSOCIATED WITH COMPONENTS SHOWN IN FIGURES 27 AND 29.[Ref. 7:p.14]

1. TAO Performance[a] affects perceived *chances of bringing about a favorable outcome to the Land Battle.*[b]

2. Ability to perform the Function (Plan, Direct, or Control) affects perceived *Close Air Support performance.*

3. Ability to perform the Function (Plan, Direct, or Control) affects perceived *Interdiction performance.*

4. Ability to perform the Function (Plan, Direct, or Control) affects perceived *Airlift performance.*

5. Features of the Elements affect perceived *ability to perform Planning for Close Air Support.*

6. Features of the Elements affect perceived *ability to perform Directing of Close Air Support.*

7. Features of the Elements affect perceived *ability to perform Controlling of Close Air Support.*

8. Features of the Elements affect perceived *ability to perform Planning for Interdiction.*

9. Features of the Elements affect perceived *ability to perform Directing of Interdiction.*

10. Features of the Elements affect perceived *ability to perform Controlling of Interdiction.*

11. Features of the Elements affect perceived *ability to perform Planning for Airlift.*

12. Features of the Elements affect perceived *ability to perform Directing of Airlift.*

13. Features of the Elements affect perceived *ability to perform Controlling of Airlift.*

[a]Independent variables (the factors to be manipulated) are underlined.
[b]Dependent variables (the response dimensions) are in italics.

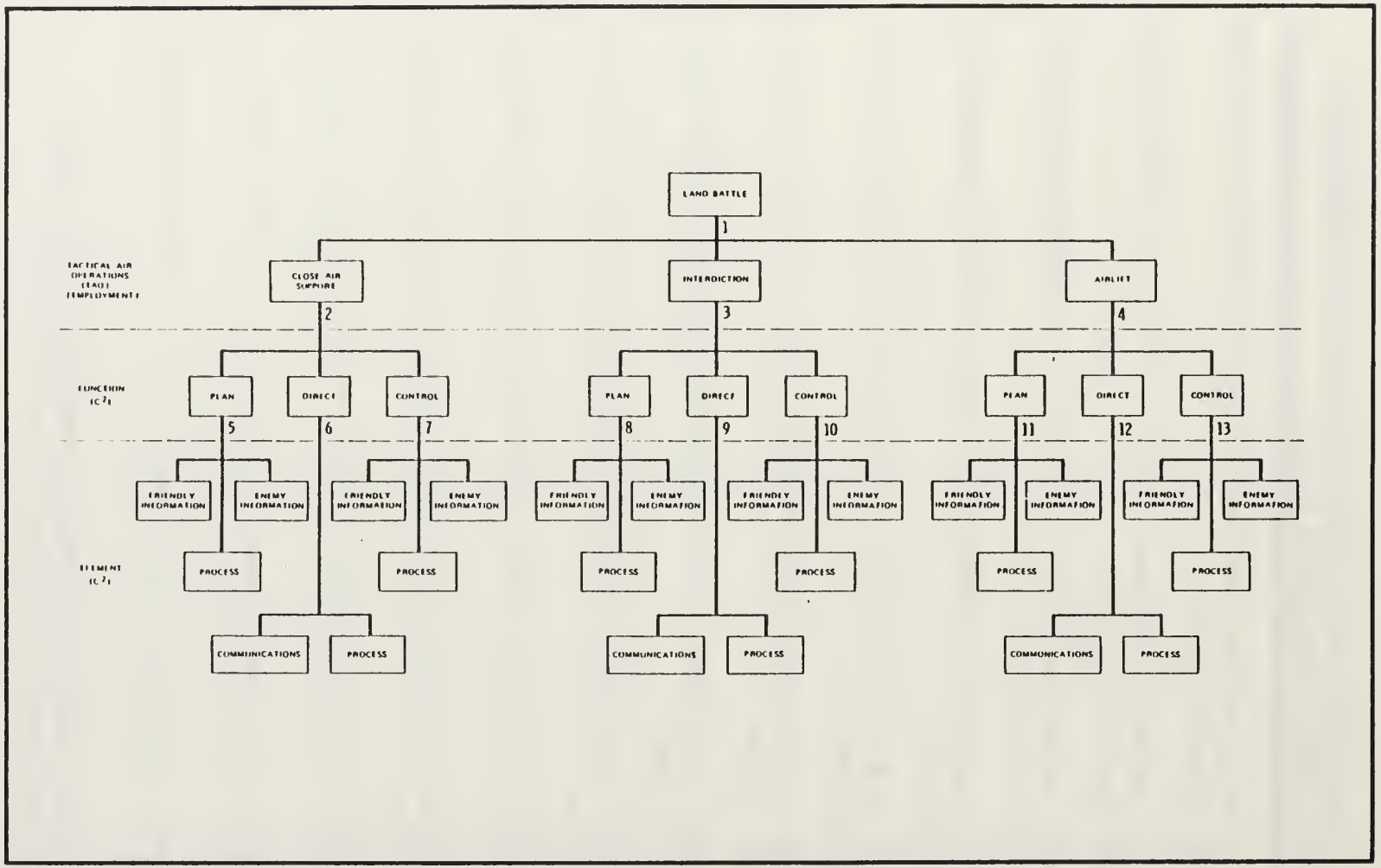

Figure 28. Tactical Air C² and Tactical Air Force Representation: Experimental Units.[Ref. 7:p.16]

One alternative structure for tactical air C^2 is provided in Figures 29A-C. Here the air/land battle is hypothesized to be affected by employing tactical air forces to perform the following tactical air actions: engage fixed targets (Figure 29A), engage stationary force elements (Figure 29B), and engage moving force elements (Figures 29C). Like the TAOs they are affected by the tactical air C^2 functions of plan, direct, and control. In addition a fourth function is determined by the expert respondents to play a role in tactical air C^2-attack capabilities of the available tactical air forces. Here the tactical air actions are system factors for the air/land battle and suboutcomes for the four tactical air C^2 functions which are factors for each of the tactical air actions.[Ref. 12:pp.5 9]

Each of the tactical air C^2 functions are suboutcomes for the following group of factors: enemy information, enemy information display, friendly information, and friendly information display. Enemy and friendly information are in turn suboutcomes for the following primitive factors: precision, amount, and currency. In addition to these, enemy information display and friendly information display are primitive factors, too.[Ref. 12:p.9]

This structure is a result of work done by The Rand Corporation as previously mentioned. Their emphasis is on relating the impact of enemy information available to the tactical air C^2 system to how it influences the impact of tactical air forces on the outcome of an airland battle. Since this thesis is concerned with relating the various levels within the evaluation hierarchy (Chapter II) to one another, the preliminary results obtained by Rand will not be discussed in further detail. It is presented here as

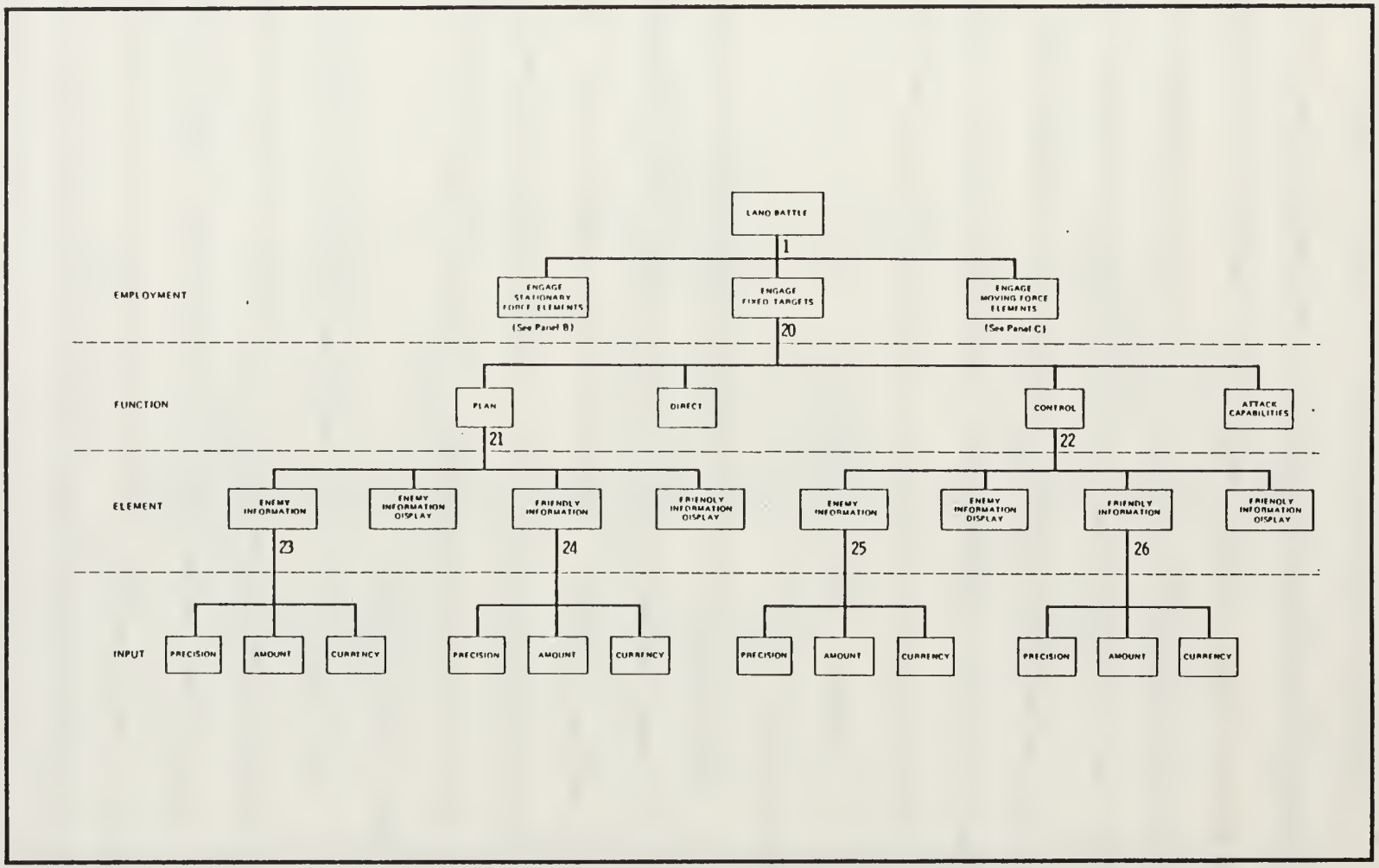

Figure 29A. Alternative Representation. Tactical Air C^2 and Tactical Air Force Representation: Fixed Targets.[Ref. 12:p.6]

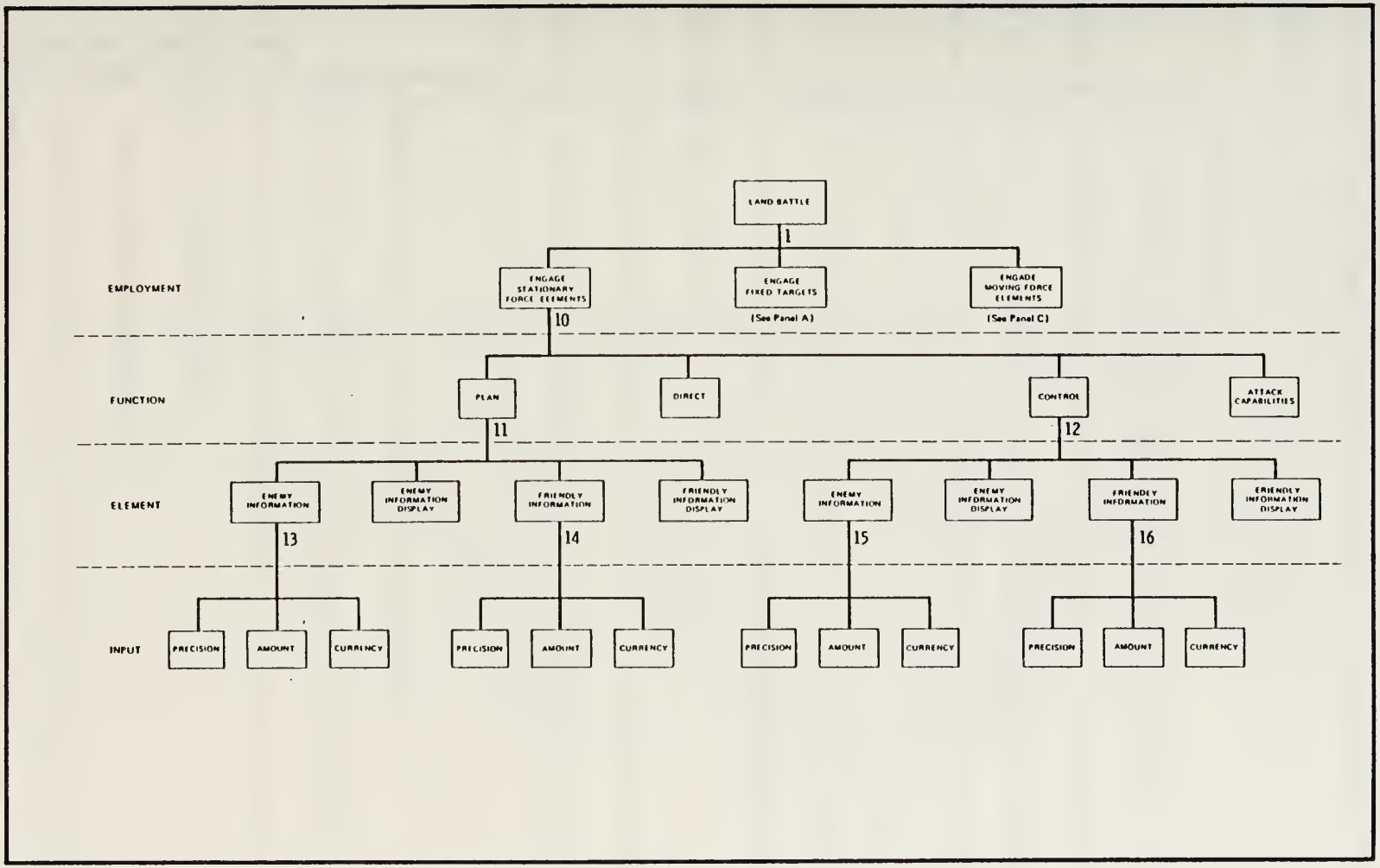

Figure 29B. Alternative Representation of Tactical Air C^2 and Tactical Air Force: Stationary Force Elements.[Ref. 12:p.7]

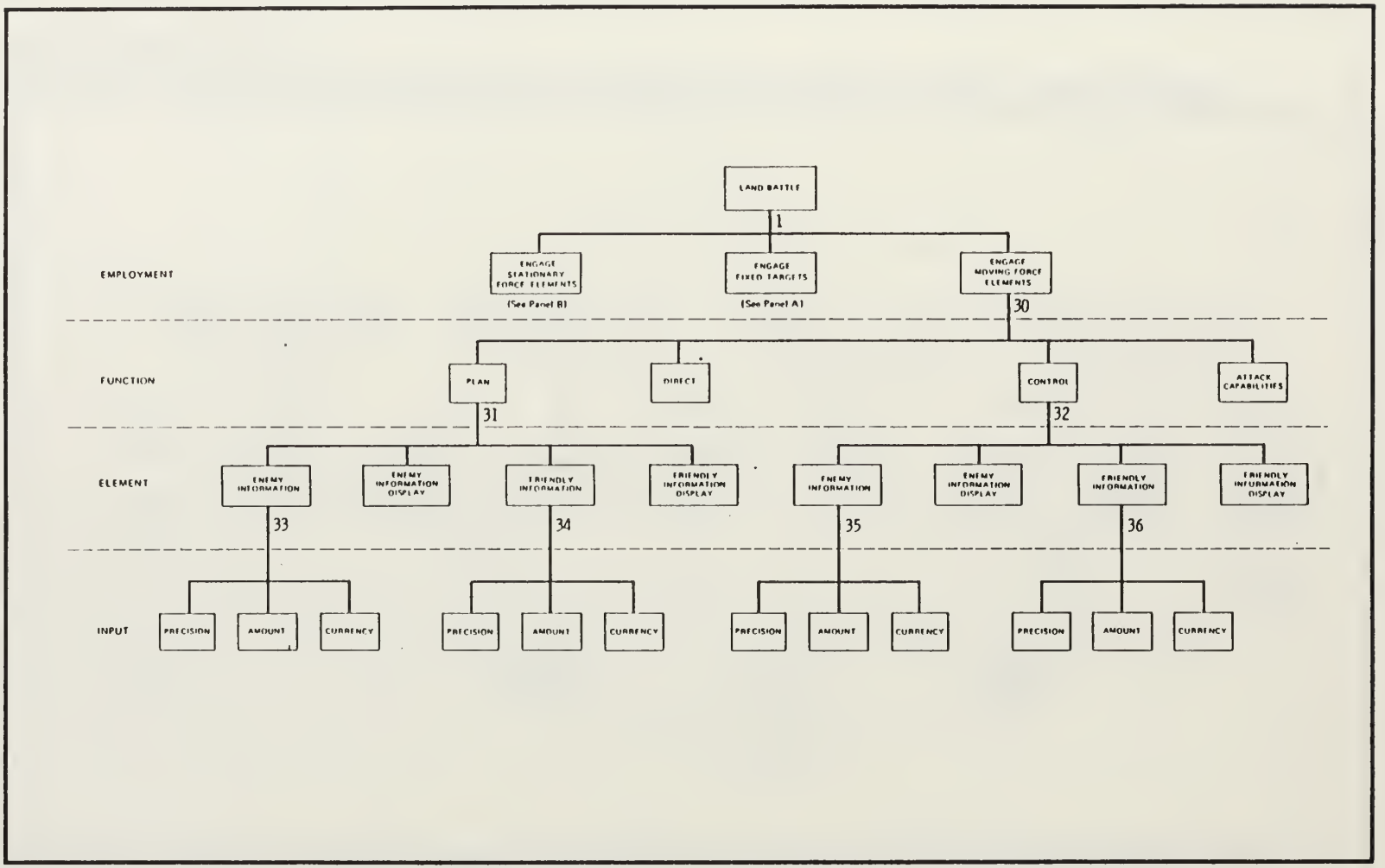

Figure 29C. Alternative Representation of Tactical Air C2 and Tactical Air Force: Moving Force Elements.[Ref. 12:p.8]

an example of an alternative structural hypothesis to the one outlined in the previous section.

2. STF Hypotheses

As discussed in Chapter IV, STF hypotheses are algebraic models, which specify the subjective processes between the perception of a stimulus and the occurrence of a response. To be considered as possible explanations of the causal relationship between a (sub)outcome and its corresponding factors, the form of the STF must be chosen in advance. This ensures adequate tests are put into the experimental design to test among their predictions. The tests of an STF will support it as an appropriate explanation when subjective values associated with the (sub)outcome and factors can be determined.[Ref. 11:pp.15-16]

Once the STF hypotheses have been determined for each experimental design, judgement experiments can be conducted for each experimental unit. The next step in the STF approach is to conduct the judgement experiments.

C. CONDUCTING THE JUDGEMENT EXPERIMENTS

In this section, hypothetical experimental designs will be discussed. Again, the focus is on showing the viability of this method and how it can be applied to the area of tactical air C^2 and tactical air forces.

1. Designing the Experiments

Assume the initial tactical air C^2 and tactical air force system is represented by Figure 28. A specific scenario (i.e. Korean conflict) would be selected as an

underlying assumption for all the experiments. An example of a single experiment performed at the TAO level (Figure 28) is outlined in TABLE IV. Here, each TAO is operationally defined in terms of its performance as being either good, fair, or poor. These factor levels are based on input from the expert respondents. A factorial design of the three factors at this tier will produce 27 questionnaire items, which would be provided to the expert respondents for judgement. Note that to adequately test the models being investigated variations on a completely crossed factorial design may be necessary. Panel C of TABLE IV details an outline of the 27 different item descriptions. Each expert respondent judges the probabilities of effecting a favorable outcome in the air\land battle based on the information in each item.[Ref. 7:pp.17-18]

Another example is provide in TABLE V. Here a hypothetical experiment for unit three at the function tier (Figure 28) is designed to test a model which specifies the effects described in the third preliminary hypothesis (TABLE III). Note, the factors, ability to plan, direct, and control can be described as good, fair, or poor. As in the previous example, a questionnaire of 27 different items would be produced from a simple factor design of all three factors. Next, a combination function (STF hypothesis) is chosen which describes the relationship between the ability to perform the functions and the interdiction performance in the air/land battle.[Ref. 7:p.18]

This same approach is applied for each of the nine units at the element level (Figure 28). Each component is described separately in terms of a certain dimension of interest. For example the components: friendly and enemy information, at unit 8 (Figure

TABLE IV. OULINE OF A POSSIBLE JUDGEMENT EXPERIMENT FOR UNIT 6 AT THE FUNCTION LEVEL.[Ref. 7:p.17]

A.	Factors		Close Air Support Performance	Interdiction Performance	Airlift Performance
B.	Factor Levels		Good Fair Poor	Good Fair Poor	Good Fair Poor
C.	Item Descriptions	1.	Good	Good	Good
		2.	Good	Good	Fair
		3.	Good	Good	Poor
		4.	.	.	.
		.	.	.	.
		.	.	.	.
		.	.	.	.
		27.	Poor	Poor	Poor
D.	Sample Item[a] (3 above)	If you knew that Close Air Support performance was good, Interdiction performance was good, and Airlift performance was poor, what would you judge your chances to be of effecting a favorable outcome in the Land Battle?			

[a]All 27 items would be randomly ordered within the questionnaire.

TABLE V. OUTLINE OF POSSIBLE JUDGEMENT EXPERIMENT AT THE TAO TIER.[Ref. 7:p.18]

A.	Factors		Ability to Plan Interdiction	Ability to Direct Interdiction	Ability to Control Interdiction
B.	Factor Levels		Good Fair Poor	Good Fair Poor	Good Fair Poor
C.	Item Descriptions	1.	Good	Good	Good
		2.	Good	Good	Fair
		3.	Good	Good	Poor
		4.	.	.	.
		.	.	.	.
		.	.	.	.
		.	.	.	.
		27.	Poor	Poor	Poor
D.	Sample Item[a] (2 above)	If you knew that the ability to Plan Interdiction was good, Direct Interdiction was good, and Control Interdiction was fair, what would you judge the Performance of Interdiction to be?			

[a]All 27 items would be randomly ordered within the questionnaire.

28) could be defined in terms of the dimension of currency. Currency is defined here to be how frequently the battle field is observed and the time it takes to get the information to the C^2 system. For the process component, the dimension of interest might be the time it takes to process incoming information. The levels of each of these factors are defined and factorially combined. This would provide the questionnaire items. For each questionnaire item, the expert respondents would be asked to judge the ability to plan interdiction. The other experimental units at the unit level would use the independent and dependent variables corresponding to their hypotheses (TABLE III).[Ref. 7:p.19]

2. Constructing Definitions of System Outcomes and System Factors.

Constructing the definitions of the system (sub)outcomes (dependent variables) and system factors (independent variables) for each judgement experiment is a critical step in the STF approach. These definitions provide the transfer feature of the combination models and allow for the experimental units to be linked functionally throughout the complex representation.[Ref. 7:p.19]

In Figures 29A-C, every component (except for those at the lowest and highest tiers) is an independent variable in one experimental unit and a dependent variable in another experimental unit. In experimental unit 5, plan is the dependent variable for the following independent variables: friendly information, process, and enemy information. But in experimental unit 2, plan is an independent variable like direct and control. The dependent variable for unit 2 is close air support. In turn, close air support is an independent variable like interdiction and airlift for the dependent variable air\land battle (land battle in Figures 29A-C). The transfer functions result from operationally

defining the components serving as both independent and dependent variables in the same terms for both experimental units, which they are a part of. This means the operational definition of close air support as a dependent variable is the same as its operational definition as an independent variable. Matching operational definitions for the other variables are given in TABLE III. When combination models are developed for the experimental units in the representation, scale values of a dependent variable (response scale value, r_{ijk} in Figure 15) in one experimental unit are defined on the same scale of values as the scale values of its associated independent variable (subjective scale value, s_i in Figure 15) at the next higher tier in the representation. Because these scale values match, the combination models specified for each experimental units are the transfer functions. By using the models as transfer functions, an output value (r_{ijk}) obtained by computing a function at one tier of the hierarchy is transferred for use as an input value for the associated model of the experimental unit at the next higher tier in the hierarchy.[Ref. 7:p.19]

The following is a look at the models which might be developed for experimental units one and two (Figure 29). A known function of the values of planning, directing, and controlling is used as a model for the variables in experimental unit two. The model is used to compute these known values to produce as an output, the value of close air support performance. A known function of the values for close air support, interdiction, and airlift performance would be used for experimental unit one. With these values, the function can be used to calculate the model's output, air\land battle (land battle in Figure 29). The input values for this model could be obtained as follows. Close

air support performance would be obtained by calculating the output to the model at unit two. For interdiction performance, the model at unit three would be used and for close air support performance, unit four would be used. Using these values, the outcome for the model at unit one would be calculated. These combination models or transfer functions are developed for each experimental units based on its respective judgement experiment. Figure 30 shows representation for tactical air C^2 and tactical air forces labeled with the transfer functions.[Ref. 7:p.19]

After the judgement experiments are designed and data for the experimental designs has been collected, the subjective transfer functions for each experimental unit would be determined. The final step of the STF approach is evaluation of the complex representation. The next section looks at evaluating the tactical air C2 and tactical air forces representation (Figure 30).

D. EVALUATING TACTICAL AIR C2 AND TACTICAL AIR FORCES

The STFs for each experimental unit (Figure 30) and the final structure are assumed to have been determined. Now the representation for tactical air C2 and tactical air forces can be evaluated. In order to evaluate this complex representation, the subjective value, r_{ijk} (Figure 15) obtained from a transfer function at one tier in the representation must be transferred to the next higher tier in the representation along the same path. For the representation in Figure 30 there are nine paths. Each path corresponds to an element group.[Ref. 7:p.22]

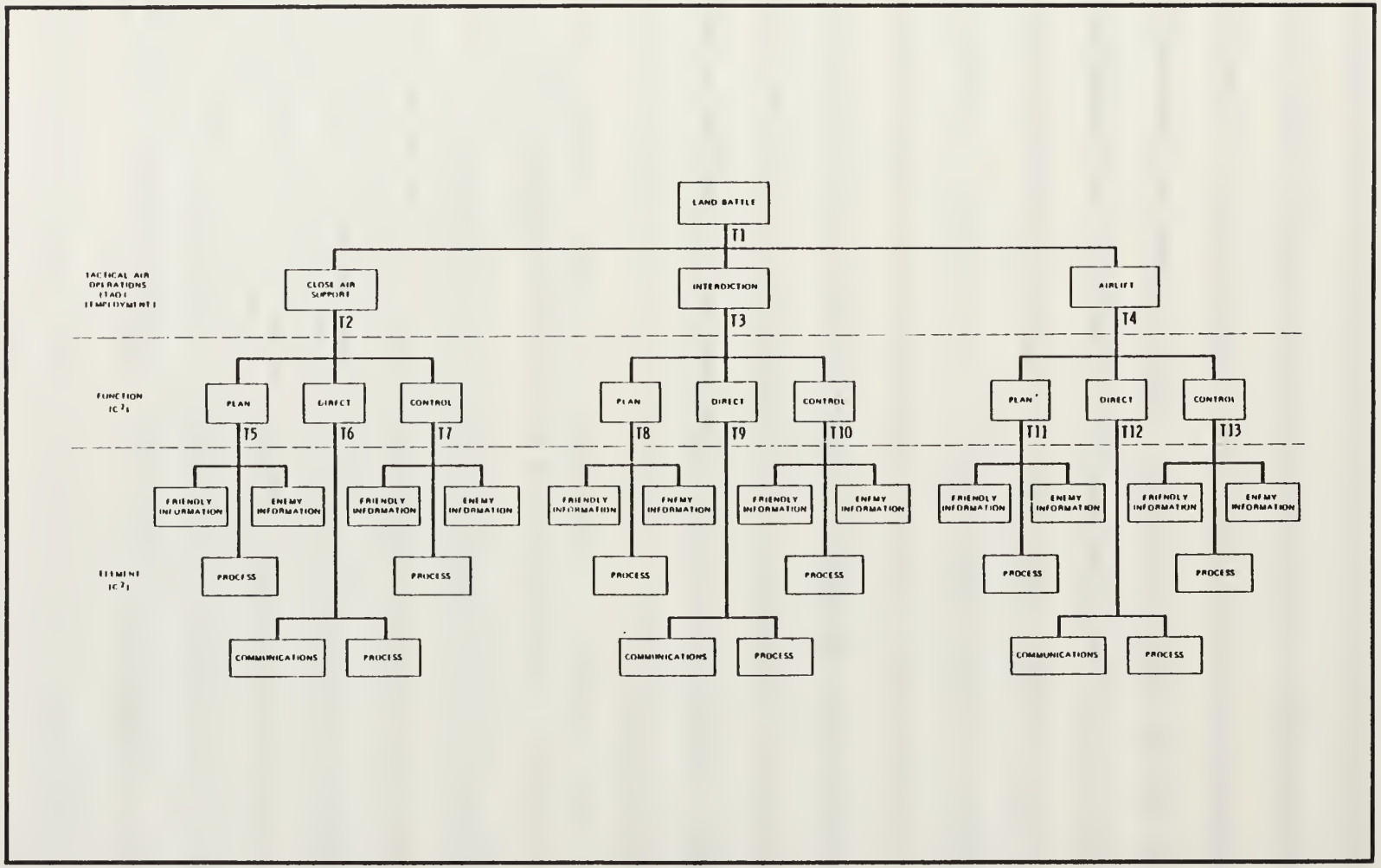

Figure 30. Tactical Air C² and Tactical Air Force Representation: Transfer Functions.[Ref. 7:p.20]

The evaluation process for the tactical air C2 and tactical air forces representation is shown in Figure 31. Input values to T5, T6, and T7 produce a subjective scale value Ψ (r_{ijk} in Figure 15) for each function. The response scale values obtained by calculating T5 is used as the scale value for planning (p), which is used to calculate T2. The response scale value, obtained by calculating T6, is used as the scale value for directing (d). This value is also needed to calculate T2. This procedure is continued until all input scale values for T2, T3, and T4 are calculated. Similarly, the response scale values obtained by calculating T2, T3, and T4 are the input scale values for T1. The resulting output obtained by calculating T1 is the overall subjective effectiveness index of tactical air C2 with respect to tactical air forces.[Ref. 7:p.22]

The subjectiveness effectiveness index provides a quantitative evaluation of the relationship between the MOE of tactical air C^2 and the MOFE of tactical air forces. The factors and/or factor levels of an experimental units in the system representation can be changed to represent other tactical air C^2 systems. A comparison of the relationship between the MOE of several tactical air C2 systems and the MOFE of tactical air forces can be done by comparing the system effectiveness indexes. Thus, transfer functions can be used to compare the outcomes within a system and the overall system outcome.[Ref. 7:p.25]

This chapter has shown how the STF approach can be applied to evaluate the relationship between tactical air C^2 and tactical air forces. The major aspects of the STF approach are the use of expert respondents and algebraic modeling. The expert

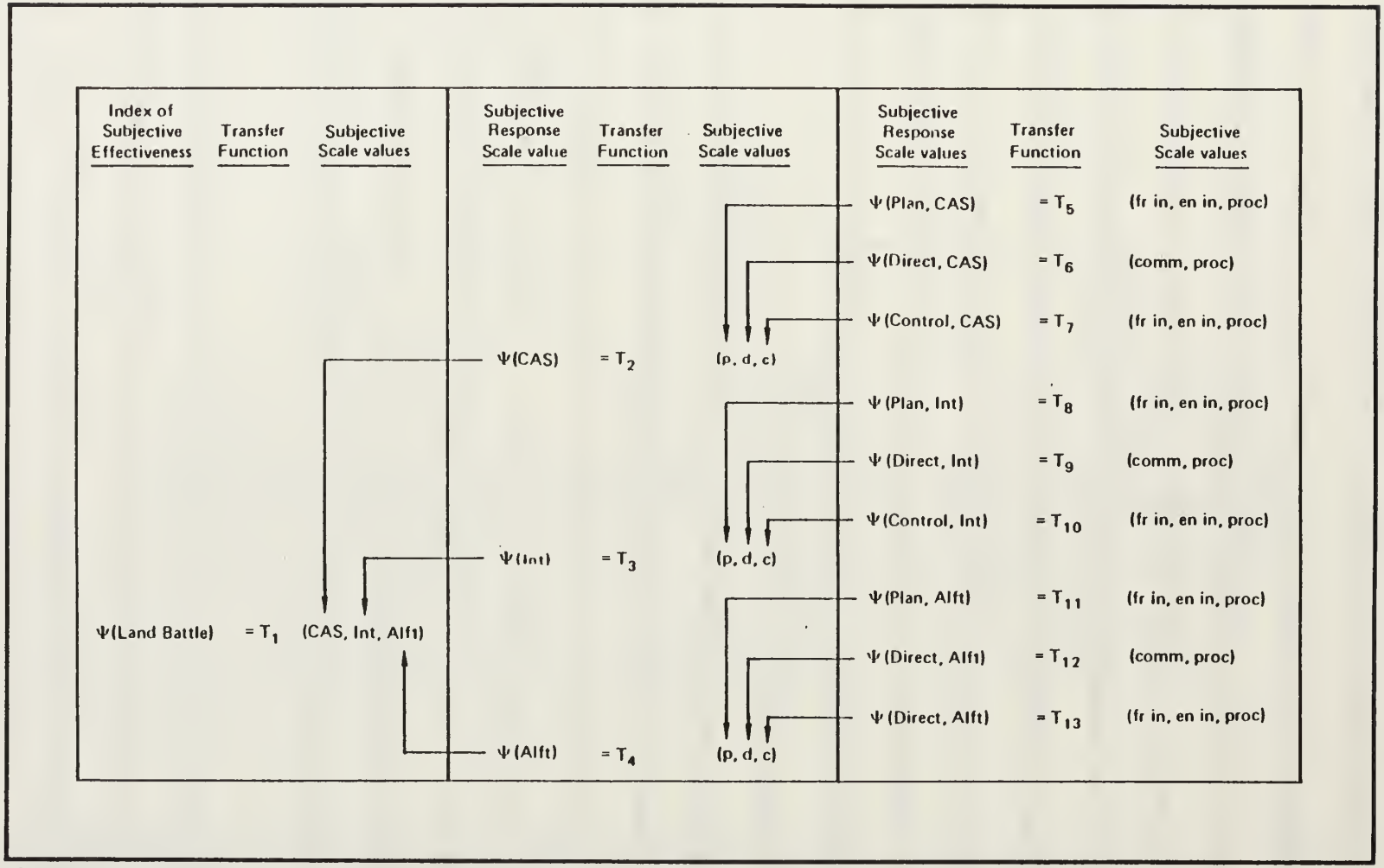
Index of Subjective Effectiveness
Transfer Function
Subjective Scale values
Subjective Response Scale value
Transfer Function
Subjective Scale values
Subjective Response Scale values
Transfer Function
Subjective Scale values
Ψ(Land Battle) = T1 (CAS, Int, Alft)
Ψ(CAS) = T2 (p, d, c)
Ψ(Int) = T3 (p, d, c)
Ψ(Alft) = T4 (p, d, c)
Ψ(Plan, CAS) = T5 (fr in, en in, proc)
Ψ(Direct, CAS) = T6 (comm, proc)
Ψ(Control, CAS) = T7 (fr in, en in, proc)
Ψ(Plan, Int) = T8 (fr in, en in, proc)
Ψ(Direct, Int) = T9 (comm, proc)
Ψ(Control, Int) = T10 (fr in, en in, proc)
Ψ(Plan, Alft) = T11 (fr in, en in, proc)
Ψ(Direct, Alft) = T12 (comm, proc)
Ψ(Direct, Alft) = T13 (fr in, en in, proc)

Figure 31. Evaluation of Tactical Air C2 and Tactical Air Force Representation.[Ref. 7:p.23]

respondents define the factors comprising a complex system and the relationship between the factors (hypotheses). The algebraic models (STFs) allow for a quantitative assessment (subjective effectiveness index) of the system. In order to now determine the usefulness of the results of the STF approach, it is important to establish its credibility. This is a difficult task at best, since the author is not aware of a widely accepted standard to compare an evaluation methodology to. Chapter VI shows how the STF approach could be evaluated. It discusses what important factors need to be considered for a meaningful and thorough evaluation to be conducted.

VI. EVALUATION CRITERIA FOR ASSESSING THE STF APPROACH

A. OVERVIEW

After applying the methodology to a problem, the next step in the scientific method process [Ref. 13:p.38] is to confirm or deny the hypothesis. In this case, it equates to evaluating the methodology. Several obstacles exist, which preclude an in depth evaluation from being conducted here. First, data was never gathered for testing the methodology, due to resource constraints on the author. Second, if the data could be obtained, no standards currently exist for verifying and validating the credibility of the methodology at this time. Yet there is a requirement for some kind of evaluation or comparison to be done.

B. EVALUATING THE APPLICATION OF THE STF APPROACH METHODOLOGY

First, a general evaluation of the STF approach methodology will concentrate on applying the six following criteria, which Erik Hollnagel uses in his book "The Reliability of Expert Systems" [Ref. 14] for evaluating expert systems:

1. correctness of the final evaluation
2. accuracy of the final evaluation
3. correctness of the reasoning techniques
4. sensitivity
5. robustness
6. cost-effectiveness.

Second, criteria for evaluating the algebraic modeling of expert responses will be outlined. Modeling is the key to the STF approach methodology and a major part of its reasoning technique. In order to determine the correctness of this reasoning technique, it is necessary to evaluate it.

1. Evaluation of the STF Approach Methodology

If the STF approach could have been fully implemented here, then the following criteria could serve as a basis for evaluating the application of the methodology. As with applying the STF approach to tactical air C^2 and tactical air forces, these evaluation criteria can only be shown how they can be applied to evaluate the STF approach.

The first criteria is the correctness of the evaluation. This looks at whether or not the output of the methodology passes the reasonability or common sense test. We expect that the MOE of tactical air C^2 impacts the MOFE of tactical air forces. If the methodology predicted no correlation between the two, then something must be wrong with the methodology.

From Chapters IV and V, the STF approach appears to meet this criteria. Experts in the system being analyzed are heavily relied upon for providing the data used in evaluating the system. However, the area which determines the impact of one component is accomplished by algebraic modeling. The skills of the researcher who incorporates the specific transfer function into the experimental design and the assumptions of the modeling process are areas which need to be further evaluated to determine their impact on the correctness of the final decision. Criteria for evaluating the

algebraic modeling will be covered later in the chapter. In order to assess this criteria the following factors need to be evaluated: the credibility of the experts, the expertise of the researchers conducting the experiments, and the assumptions of the algebraic models.

The second criteria is the accuracy of the evaluation. The degree to which the evaluation is correct is useful only if the evaluation passes the first criteria. What is of interest here is in one sense the reasonability of the accuracy of the evaluation. This can be verified by applying the methodology to a simple system with a known evaluation. In other words, the methodology which is designed to analyze complex systems should provide the same evaluation of the simple system. Another approach to assessing the accuracy of the evaluation is called the Delphi Method [Ref. 15]. The Delphi Method was developed by The Rand Corporation for problem analysis when little hard data is available on the problem or system being analyzed. A group of carefully selected experts knowledgeable on the system arrive at consensus on answers to questions about the performance of the system under specific conditions. Because face-to-face group discussion of the questions is not allowed, group think and other disadvantages of group discussion are avoided. The key to the Delphi Method is using expert opinion to verify the accuracy of the evaluation. In order to assess this criteria the following factors need to be evaluated: the credibility of the expert respondents, the expertise of the researchers in conducting the experiment, the assumptions of the algebraic models, and the validity of the evaluation.

The third criteria is the correctness of the reasoning techniques. It asks whether or not the logic in the methodology is consistent with the output. For example,

if we are going to make a cake we don't follow a recipe for a pie. In the STF approach, the algebraic modeling used to determine the STF is the major factor of the reasoning technique. The other factor is the researchers designing the judgement experiments and analyzing the data to come up with the correct STFs. In assessing this criteria the following factors need to evaluated: the effectiveness of the algebraic models and the construction of the models.

The fourth criteria is sensitivity. Sensitivity is defined as the minimum difference in input required to affect the outcome of the evaluation [Ref. 14:p.177]. The concern is to guard against the methodology acting as a filter, ensuring that only real changes to the input affect the output. The STF approach ensures this by having the experts define the factors and the factors levels. In order to assess this criteria the following factors needs to be evaluated: the credibility of the expert respondents and the construction of the algebraic models.

The fifth criteria is robustness. This is the ability to handle the nonstandard input. In a sense, it is a selective kind of sensitivity. If the methodology is robust then it ensures it is sensitive to certain inputs and insensitive to others. Problems can arise in determining the robustness if it is difficult to determine all the different types of nonstandard input. Another perspective of robustness is that it prevents the old adage, garbage-in garbage-out. A robust methodology ensures that garbage-in results in nothing out. For the STF approach, it requires analyzing the algebraic modeling and knowing how to handle expert responses which do not fit any of the standard combination models (interactive and noninteractive functions) used as STFs. In order to assess this criteria

the following areas need to be evaluated: the expertise of the researchers in conducting the experiment, the effectiveness of the algebraic models, and the construction of the algebraic models.[Ref. 14:pp,177-178]

The sixth criteria is cost-effectiveness. Cost-effectiveness is defined here as the comparison of benefits of applying the methodology to the overall cost (manpower, resources, etc.) involved. From the general application of the STF approach it appears to be very time consuming for the experts involved in the judgement experiments and even more so for the researchers conducting the STF approach. However, from a cost perspective, it may be more cost-effective than other methodologies providing the same type of evaluation. In order to assess this criteria the following areas need to be evaluated: the initial cost of competing methodologies, the recurring cost of competing methodologies, the expertise of the researchers, the credibility of the experts, and the significance of the statistics.[Ref. 14:p.179]

The TABLE VI summarizes the assessment factors for each of Hollnagel's criteria applied to the STF approach and TABLE VII summarizes the overall assessment of these criteria. There are two limiting factors in evaluating this methodology. First, neither an evaluation of the methodology nor an evaluation of the algebraic modeling was done in full. Second, the criteria being used to evaluate the potential of the STF approach is not a universal or even locally accepted standard. The second limiting factor can be waived by author's prerogative as this is what seemed to reasonably cover all the major aspects of a methodology. This leaves the first limiting factor as the overall limiting factor.

TABLE VI. SUMMARY OF ASSESSMENT FACTORS.

EVALUATION CRITERIA[1]	RELEVANT FEATURES	ASSESSMENT FACTORS
CORRECTNESS OF THE FINAL EVALUATION.	•Expert responses to judgement questionnaires.	•Credibility of experts.
	•Algebraic models of expert response.	•Expertise of researchers in conducting the experiment. •Assumptions of the algebraic models.[2]
ACCURACY OF THE FINAL EVALUATION.	•Expert responses to judgement questionnaires.	•Credibility of experts.
	•Algebraic models of expert response.	•Expertise of researchers in conducting the experiment. •Validity of evaluation: ••Apply to system with a known evaluation. ••Delphi Method.
CORRECTNESS OF THE REASONING TECHNIQUES.	•Expert responses to judgement questionnaires.	•Expertise of researchers in conducting the experiment.
	•Algebraic models of expert response.	•Effectiveness of the algebraic models.[2] •Construction of the algebraic models.[2]

TABLE VI. (Continued) SUMMARY OF ASSESSMENT FACTORS.

EVALUATION CRITERIA[1]	RELEVANT FEATURES	ASSESSMENT FACTORS
SENSITIVITY.	•Expert definition of factor and factor levels.	•Credibility of experts.
ROBUSTNESS.	•Algebraic models of expert response.	•Expertise of researchers in conducting the experiment. •Effectiveness of the algebraic models.[2] •Construction of the algebraic models.[2]
COST-EFFECTIVENESS.	•Initial cost: ••Developing the methodology.	•Initial cost of competing methodologies.
	•Recurring cost: ••Salaries of researchers. ••Travel costs.	•Recurring cost of competing methodologies.
	•Payoff in terms of Hollnagel's six criteria.	•Expertise of researchers in conducting the experiment. •Credibility of experts. •Statistical significance of the system effectiveness index.

[1]Hollnagel's criteria for evaluating expert systems.
[2]Giordano and Weir's criteria for evaluating models.

TABLE VII. SUMMARY OF OVERALL ASSESSMENT OF EVALUATION CRITERIA.

EVALUATION CRITERIA	OVERALL ASSESSMENT
CORRECTNESS OF THE FINAL EVALUATION.	The use of expert reasoning and algebraic modeling is a valid technique.
ACCURACY OF THE FINAL EVALUATION.	Given qualified experts and valid modeling, the accuracy of the STF approach should be excellent.
CORRECTNESS OF THE REASONING TECHNIQUES.	The correctness of the reasoning techniques used in the STF approach is strongly a function of the model/ experiment designer. This requires a high degree of technical/military competence which may limit the widespread use of the approach.
SENSITIVITY.	The STF approach provides a means for sensitivity analysis.
ROBUSTNESS.	The STF approacah is robust in that the model/experiment definitions are tailored to the specific application.
COST-EFFECTIVENESS.	The STF approach is very complex to implement and requires the time consuming process of soliciting expert responses. While it may be costly, it may be the only technique available to make an adequate evaluation of a large complex system.

Without an in depth evaluation of the algebraic modeling, the accuracy of the final evaluation, the correctness of the reasoning technique and the sensitivity and robustness of the methodology can not be fully evaluated with respect to the STF approach. The cost-effectiveness of the STF approach will depend on initial costs and recurring costs of applying the methodology. It appears that the recurring costs will always comprise the majority of the costs, since salaries, travel costs will always be incurred. The initial cost of this methodology appears to be small, because it doesn't require the extensive computing power of (for example) a methodology which relies on a computer simulation. In contrast, if a competing methodology uses a computer simulation then initial costs would appear to outweigh recurring costs. This is assuming initial costs would probably include developing the computer code for the simulation and procuring the equipment and software to run the simulation on and recurring costs for obtaining data for the simulation. Again, the cost-effectiveness cannot be evaluated in a meaningful manner as other methodologies were not available for comparison. Thus, as this approach is so aptly named, this critique of the STF approach is purely subjective.

2. Evaluation of Algebraic Modeling of Expert Responses

The heart of the STF approach is the algebraic modeling used for determining the STFs. To fully evaluate the STF approach an in depth analysis and evaluation of the modeling process is required. The following criteria for evaluating the algebraic modeling is presented as a viable approach to help determine the validity of the STF approach by evaluating the validity of the STFs.

In their textbook "A First Course in Mathematical Modeling", Frank R. Giordano and Maurice D. Weir provide guidelines for evaluating the properties of a model and for evaluating the construction of a model.[Ref. 13]

a. Evaluation of the Effectiveness of a Model

Giordano and Weir list the three following properties of a model as important in determining the capabilities of a model:

1. fidelity
2. cost
3. flexibility [Ref. 13:p.33].

An evaluation of the effectiveness of the model in terms of these capabilities can be done. This helps to determine if the model chosen is suited to the system it is being used to model.

Fidelity is the precision with which a model represents reality [Ref. 13:p.33]. In the STF approach, the algebraic models are used to represent the response of the experts to the questionnaire items.

The second property is cost. This is the total cost of the modeling process [Ref. 13:p.33]. Related to the cost of the model is the cost-effectiveness of the model. Similar to determining the cost-effectiveness of a methodology, the cost-effectiveness of a model needs to be taken into consideration. Since, in this case the cost-effectiveness of the STF approach methodology is dependent on the cost-effectiveness of the algebraic modeling.

The last property is flexibility. This pertains to the ability of the model to change and control conditions, which affect the model as the data is being collected

[Ref. 13:p.33]. This is similar to the robustness of the methodology in terms of how well it can handle data, which is not accurately represented by the more commonly used models. The ability to model any pattern of expert responses is required for the algebraic models used in the STF approach.

In terms of the types of models required for the STF approach, all of these are important in choosing the correct models to use for the STFs. After choosing a model which meets the requirements required for fidelity, cost, and flexibility, the construction of the model should be evaluated.

b. Evaluation of the Construction of a Model

Giordano and Weir define the methodology for constructing a mathematical model as consisting of the following five steps:

Step 1. Identify the problem.
Step 2. Make assumptions.
  a. Identify and classify the variables.
  b. Determine interrelationships between the variables and submodels.
Step 3. Solve the model.
Step 4. Verify the model.
  a. Does it address the problem?
  b. Does it make common sense?
  c. Test it with real-world data.
Step 5. Implement the model.
Step 6. Maintain the model.[Ref. 13:p.37]

By evaluating each step within the methodology, the construction of the model can be evaluated. This area requires that the individual evaluating the model being used is well versed in the art of modeling. Modeling requires more than a "cookbook" of the different

types of models. It requires a "grab bag" of models, which are the exceptions to the rule and whose use usually comes through experience alone.

Thus, the evaluation of the modeling process used in the STF approach should be left to an expert in that area. The two evaluations (Giordano and Weir's criteria for evaluating the effectiveness of a model and for evaluating the construction of a model) outlined above provide an overall picture of what an expert modeler might concentrate on when evaluating a model.

Specific conclusions about the relationship between the MOE of a tactical air C^2 and the MOFE of a tactical air force are not appropriate at this time. Further, an in depth evaluation of the STF approach cannot be accomplished for the lack of data and more importantly the lack of standard criteria to be evaluated against. This chapter showed how Hollnagel's criteria for evaluating the STF approach methodology and Giordano and Weir's criteria for evaluating the modeling used in this methodology could be applied. Regardless of whether the STF approach is used or another methodology is used to evaluate the relationship between MOEVs, the methodology used should act solely as a guide. Other factors which the methodology may not include need to be taken into consideration.

Sometimes when trying to solve a problem the answer is not found after the problem has been analyzed and broken down into its basic components. What is discovered is that some of the basic components and their relationships are missing or have not been adequately defined yet. This is not to say that a solution to the problem cannot be found. It just cannot be found until the other components have been identified

and/or defined. The last chapter makes recommendations as to those basic components and their relationships, which the author feels need to be addressed before the solution to the problem (evaluating the relationship between the MOEVs) can be found.

VII. SUMMARY AND RECOMMENDATIONS

A. OVERVIEW

This chapter presents a summary of the thesis and recommendations for improving the evaluation of combat effectiveness. The key points will first be summarized. Finally, several key issues and recommendations are made in the following areas: 1) the defense planning structure, 2) the military system hierarchy, 3) the definitions of MOEVs, 4) the hierarchical relationship among MOEVs, and 5) the assessment of evaluation methodologies.

B. SUMMARY

It has been said that the first step in solving a problem is to break it down into its basic components and establish the relationships existing between them. This provides a strong foundation for solving the problem. In Chapter I, the general problem of controlling combat effectiveness was identified and served as a motivator for evaluating the relationship between various MOEVs. In particular, the problem of evaluating the contribution of tactical air C^2 to tactical air forces was chosen. The thesis focused on one methodology for evaluating this relationship.

In breaking a problem into its basic components, the components should be defined. Chapter II defined the basic MOEVs, which the author felt comprise a hierarchy of MOEVs. A broader view of what is being evaluated is required before the MOEVs can

be defined. One possible structure for the defense planning framework developed by The Rand Corporation was presented. This framework determines the structure of the military system. A possible hierarchical structure for the military system was presented. It provided the link between the defense planning structure and the hierarchy of MOEVs. The MOEVs can now be defined to correspond to each level of the military system. Tactical air C^2 and tactical air forces can be viewed as two components (system and force) within the military system. The measures of evaluation for these two components can be viewed as two components (MOE and MOFE) within the hierarchy of MOEVs. Thus, a partial foundation had been established for evaluating the relationship between tactical air C^2 and tactical air forces. Two key relationships had to be addressed: 1) the basic components of tactical air C^2 and tactical air forces and their relationships and 2) a methodology for evaluating the relationship between MOEVs, in this case the relationship between the MOE of tactical air C^2 and the MOFE of tactical air forces.

Chapter III described the components for tactical air C^2 and tactical air forces and the causal relationships between them. It established the MOE for tactical air C^2 and tactical air forces. It started out by looking at the overall picture of combat, the air/land battle. The relationship between the air/land battle and tactical air forces was established. Next, the relationship between tactical air C^2 and tactical air forces is defined. The emphasis was on identifying the basic components and their relationships.

Chapter IV described one methodology for evaluating the relationship between MOEVs in the MOEV hierarchy: the subjective transfer function (STF) approach. The steps of this methodology are outlined based on the notes published by Rand. Finally,

with the foundation completed, the relationship between MOE of tactical air C^2 and MOFE of tactical air forces can be evaluated by applying the STF approach. Chapter V looks at one way of applying the STF approach to this relationship. The full application of this methodology is beyond the resources of the author. However, the STF approach is shown how it could be used to evaluate the relationship between a MOE of a specific system and the MOFE of a specific force. This partial application highlights some of the advantages and disadvantages of this methodology. A general evaluation of the methodology is then possible.

Chapter VI showed one way of evaluating the STF approach. It divided the evaluation into two areas: the methodology and algebraic modeling. Erik Hollnagel's criteria was defined and used to show what factors of the STF approach needed to be evaluated. The main reasoning technique used in the STF approach is algebraic modeling. Giordano and Weir's guidelines for evaluating the effectiveness of a model and the construction of a model were presented as viable evaluation criteria for determining the credibility of the algebraic modeling used in the STF approach.

One of the reasons a full and meaningful evaluation of the STF approach methodology could not be done was a number of basic components for evaluating the relationship between MOEVs and for evaluating a methodology need to be defined and standardized. Although Chapter VI provided criteria for evaluating the STF approach, there is no universally accepted standards by which individuals or organizations assess evaluation methodologies. The next section looks at (1) key issues which make evaluating the relationship among MOEVs difficult and (2) recommendations regarding

standardization of the definitions and hierarchy of MOEVs, and the methodology used to evaluate the relationship.

C. KEY ISSUES AND RECOMMENDATIONS

The following key issues and recommendations are presented as items which need to be addressed. The key issues are presented as "facts of life" which inevitably make the process of once and for all defining MOEV relationship impossible. The recommendations, while are worth striving for, will ensure a sound foundation for defining and validating a methodology for evaluating the relationship among MOEVs.

1. Issue No. 1: The Defense Planning Framework

The defense planning framework determines the structure of the military system. In Chapter II, Glenn A. Kent's framework for defense planning [Ref. 4] was presented as one possible structure (Figure 1A) for the defense planning process. As it is constantly changing, the military system tends to be in a state of change. Thus, the relationship between the different levels in the military system may be changing. Determining a methodology to evaluate the relationship between the MOEVs requires the realization that the relationship between components in the military system will not be constant.

2. Issue No. 2: The Military System Hierarchy

In Chapter II, one possible hierarchy for the military system was presented (Figure 1B). However, the military system hierarchy is a logical outcome of the defense planning framework and as such, will also be changing. If the structure of the military

system is constantly in a state of change then so will be the relationship between the components. MOEVs and their hierarchy must be adjusted accordingly in any specific application, to ensure that the evaluation methodology yields a true picture of force effectiveness.

3. Recommendation No. 1: Standardize Definitions of MOEVs

Even if the defense planning framework and military system hierarchy are in a continuing state of flux, standardizing the definitions of MOEVs is essential. Evaluations are worthless unless there is a standard agreed upon, which serves as a basis for a meaningful comparison. Imagine that our weights and measures system were not standardized in the U.S.. A gallon of gas from one station might not be the same quantity of gas as a gallon from another station. Trying to figure out which gas station was the least expensive would not be as easy as checking the price per gallon. Without a standard measure it is difficult to realize the value in comparing items based on these measures. Whether it is the price of gas or the relationship between two MOEVs, a consensus is required on what standard will be used. As was pointed out in Chapter II, document searches for MOEs were deemed useless, since what one person defined as a MOE was what someone else defined as a MOP. Chapter II provided one set of MOEVs and their definitions. Unless MOEVS and their definitions are standardized by the Department of Defense, the deception of meaningful evaluations will continue to flourish.

4. Recommendation No. 2: Standardize MOEV Hierarchy

Essential to evaluating the relationship between MOEVs is a standardized MOEV hierarchy. This ensures at the lowest level that the relationship between MOEVs will remain constant. Chapter II provided one hierarchy for MOEVs (Figure 2B). Determining a methodology to evaluate the relationship between MOEVs demands a standard for assessing evaluation methodologies.

5. Recommendation No. 3: Develop A Standard for Assessing Evaluation Methodologies

All the basic components for developing an evaluation methodology have been addressed. This last item is necessary to verify that the methodology chosen is the best one for the job. This comes down to what attributes are important in evaluating an evaluation methodology. In general, they should be those attributes, which are desired in any methodology. Hollnagel's attributes, as discussed in the Chapter VI, were used for evaluating expert systems. Presently, they are just as good as any other set of attributes as long as the person using them can justify their usefulness. Chapter VI provided one approach to assessing the credibility of an evaluation methodology. Until a standard is developed, comparison of evaluation methodologies will be difficult at best.

Evaluating combat effectiveness like evaluating the MOE of tactical air C^2 systems requires a look at the big picture. It requires that a system's basic components and their relationships are established. Standards must then be developed and accepted as the basis for any evaluations done for that system.

LIST OF REFERENCES

1. Systems Exploration, Inc., *Tactical Command System (TCS) Measures of Effectiveness and Performance*, February 1989.

2. F.A. Tillman, C.L. Hwang, Way Kuo, *System Effectiveness Models: An Annotated Bibliography*, IEEE Transactions of Reliability, Vol. R-29, No. 4, October 1980.

3. NMCS/WWMCSS Engineering Integration Office, *Volume 4 Derivation and Application of Measures of Effectiveness (Draft)*, International Business Machines Corp., June 1978.

4. Kent, Glenn A., *A Framework For Defense Planning*, The Rand Corporation, August 1989.

5. Cushman, John H., Air/Land Battle Mastery and C^2 Systems For the Multinational Field Commanders, *Tactical C^3 For the Ground Forces*, AFCEA International Press, 1986.

6. Callero, Monti, Willard Naslund, Clairice T. Veit, *Subjective Measurement of Tactical Air Command and Control--Vol. I: Background and Approach*, The Rand Corporation, March 1981.

7. Veit, Clairice T. and Monti Callero, *Subjective Transfer Function Approach to Complex System Analysis*, The Rand Corporation, March 1981.

8. Wohl, J. G., D. Gootkind, and H. D'Angelo, *Measures of Merit for Command and Control*, The MITRE Corporation, January 1981.

9. Wohl, Joseph G., *Force Management Decision Requirements for Air Force Tactical Command and Control*, IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-11, No. 9, September 1981.

10. Hayward, Philip, *The Measurement of Combat Effectiveness*, The Journal of Operations Research Society of America, Vol. 16, No. 2, March-April 1968.

11. Veit, Clairice T., Monti Callero, Barbara J. Rose, *Introduction to the Subjective Transfer Function Approach to Analyzing Systems*, The Rand Corporation, July 1984.

12. Callero, Monti, Willard Naslund, and Clairice T. Veit, *Subjective Measurement of Tactical Air Command and Control--Vol. II: The Initial Representation*, The Rand Corporation, March 1981.

13. Giordano, Frank R. and Maurice D. Weir, *A First Course in Mathematical Modeling*, Brooks/Cole Publishing Company, 1985.

14. Hollnagel, Erik, *The Reliability of Expert Systems*, Ellis Horwood Limited, 1989.

15. Hoover, Stewart V. and Ronald F. Perry, *Simulation: A Problem-Solving Approach*, Addison-Wesley Publishing Company, Inc., 1989.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
   Cameron Station
   Alexandria, VA 22304-6145

2. Library, Code 52 2
   Naval Postgraduate School
   Monterey, CA 93943-5000

3. Director for Command, Control, and
   Communications Systems, Joint Staff 1
   Washington, DC 20318-6000

4. C3 Academic Group, Code CC 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

5. AFIT/NR 1
   Wright-Patterson AFB, OH 45433_6583

6. AFIT/CIRK 1
   Wright-Patterson AFB, OH 45433_6583

7. Professor Donald A. Lacer, Code CC/LA 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

8. Dr. Michael G. Sovereign, Code OR/SM 3
   Naval Postgraduate School
   Monterey, CA 93943-5000

9. Major Marvin A. Knorr, Jr., USMC 1
   C2 Development Branch
   C2GO
   MCRDC
   Quantico, VA 22134

10. Captain Stephen C. Kessner, USAF.....2
P.O. Box #1000
Colorado Springs, CO 80914-5001

Thesis
K39477 Kessner
c.1 A methodology for evaluating the relationship between measures of evaluation (MOEVs).

Thesis
K39477 Kessner
c.1 A methodology for evaluating the relationship between measures of evaluation (MOEVs).

DUDLEY KNOX LIBRARY
3 2768 00016530 2